

**MENTAL FUNCTION AND CEREBRAL CARTOGRAPHY:  
FUNCTIONAL LOCALIZATION IN fMRI RESEARCH**

by

**Joseph Brendan McCaffrey**

B.A. in Biology, The Colorado College, 2007

Submitted to the Graduate Faculty of

The Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH  
KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Joseph Brendan McCaffrey

It was defended on

May 4, 2016

and approved by

David Plaut, Professor of Psychology

David Danks, Professor of Philosophy

Mazviita Chirimuuta, Assistant Professor of History and Philosophy of Science

Dissertation Advisor: Edouard Machery, Distinguished Professor of History and Philosophy  
of Science

Co-Advisor: Peter Machamer, Professor of History and Philosophy of Science

Copyright © by Joseph Brendan McCaffrey

2016

# **MENTAL FUNCTION AND CEREBRAL CARTOGRAPHY: FUNCTIONAL LOCALIZATION IN fMRI RESEARCH**

Joseph Brendan McCaffrey, Ph.D.

University of Pittsburgh, 2016

This dissertation advances a novel philosophical account of the relationship between brain mapping and cognitive theorizing in functional magnetic resonance imaging (fMRI) research. I argue that testing hypotheses about human cognition and behavior with fMRI critically depends on bridging assumptions about how cognitive functions map onto the brain. I demonstrate that in light of recent theoretical (e.g., network thinking) and methodological (e.g., resting state fMRI) advancements, these bridging assumptions are often problematic. I conclude that at this stage of scientific development, fMRI research should focus on articulating and testing new bridging assumptions rather than testing psychological theories.

## TABLE OF CONTENTS

<a href="#">PREFACE.....</a>	<a href="#">ix</a>
<a href="#">1.0 DISSERTATION INTRODUCTION.....</a>	<a href="#">1</a>
<a href="#">1.1 PROJECT DESCRIPTION.....</a>	<a href="#">1</a>
<a href="#">1.2 HISTORICAL AND PHILOSOPHICAL BACKGROUND.....</a>	<a href="#">4</a>
<a href="#">1.3 CONTROVERSIES ABOUT fMRI.....</a>	<a href="#">6</a>
<a href="#">1.4 OUTLINE OF DISSERTATION.....</a>	<a href="#">10</a>
<a href="#">2.0 THE BRAIN'S HETEROGENEOUS FUNCTIONAL LANDSCAPE.....</a>	<a href="#">14</a>
<a href="#">2.1 CHAPTER 2 INTRODUCTION.....</a>	<a href="#">14</a>
<a href="#">2.2 THE PROBLEM OF MULTI-FUNCTIONALITY.....</a>	<a href="#">16</a>
<a href="#">2.3 FUNCTIONAL MAPPING STRATEGIES.....</a>	<a href="#">19</a>
<a href="#">2.3.1 Subdivide and conquer.....</a>	<a href="#">20</a>
<a href="#">2.3.2 Cognitive ontology revision: new systematic mappings.....</a>	<a href="#">21</a>
<a href="#">2.3.3 Context-sensitive functional mappings.....</a>	<a href="#">23</a>
<a href="#">2.4 DIFFERENT PARTS, DIFFERENT STRATEGIES.....</a>	<a href="#">25</a>
<a href="#">2.4.1 Mechanistic organization and functional mappings.....</a>	<a href="#">26</a>
<a href="#">2.4.2 Roles, capacities, and kinds of multi-functional parts.....</a>	<a href="#">27</a>
<a href="#">2.4.3 Conserved roles and systematic mappings.....</a>	<a href="#">30</a>
<a href="#">2.4.4 Variable roles and context-sensitive mappings.....</a>	<a href="#">33</a>

2.5 THE FUNCTIONAL HETEROGENEITY HYPOTHESIS.....	36
3.0 COGNITIVE ONTOLOGY REVISION AND fMRI: TWO STRATEGIES.....	39
3.1 CHAPTER 3 INTRODUCTION.....	39
3.2 COGNITIVE ONTOLOGY REVISION.....	43
3.3 fMRI AND THE CASE FOR REVISION.....	48
3.3.1 The call for revision.....	48
3.3.2 Evidence for revision.....	50
3.3.3 Bridging assumptions in fMRI studies.....	53
3.4 A BRIDGE TOO FAR? FAILURES OF BRIDGING ASSUMPTIONS.....	55
3.4.1 Neural reuse and dedicated substrates.....	55
3.4.2 A case: fMRI and basic emotion theory.....	60
3.5 PATTERN CLASSIFICATION AND COGNITIVE ONTOLOGIES.....	64
3.5.1 Classifier performance and cognitive constructs.....	66
3.5.2 Two confounds: additional constructs and implementation differences....	70
3.5.3 Controlling for confounds.....	74
3.6 RETHINKING THE DIRECTION OF REVISION.....	75
4.0 NETWORK INFERENCES AND THE USES OF RESTING STATE fMRI.....	78
4.1 CHAPTER 4 INTRODUCTION.....	78
4.2 THE USES OF RESTING STATE fMRI.....	84
4.2.1 Resting state analyses.....	85
4.2.2 Interpretations of resting state data.....	88
4.3 THE CHALLENGE OF MIXTURE DISTRIBUTIONS.....	92
4.3.1 Functional networks or sampling artifacts?.....	92

4.3.2 Sampling a mixture distribution.....	94
4.3.3 More evidence for the mixture view.....	99
4.4. OBJECTIONS AND REPLIES.....	101
4.4.1 Objection 1: consistency of resting state functional connectivity patterns.....	102
4.4.2 Objection 2: timing and magnitude of resting state fluctuations.....	104
4.4.3 Objection 3: resting state fluctuations persist without consciousness.....	105
4.5 THEORETICAL AND METHODOLOGICAL CONSEQUENCES.....	107
4.5.1 New tools and techniques.....	107
4.5.2 The value of resting state data.....	110
4.6 EXPLORING BRAIN NETWORKS.....	112
5.0 DISSERTATION CONCLUSION.....	114
5.1 THE STORY SO FAR.....	114
5.2 BRAIN MAPPING: HOW BAD COULD IT BE?.....	115
5.3 COGNITIVE ONTOLOGY REVISION: A ROADMAP.....	117
5.4 MENTAL FUNCTION AND CEREBRAL CARTOGRAPHY.....	118
BIBLIOGRAPHY.....	119

## LIST OF FIGURES

<a href="#">Figure 2.4 Roles versus Capacities in Human Brain Mapping.....</a>	<a href="#">30</a>
<a href="#">Figure 2.5 The Functional Heterogeneity Hypothesis.....</a>	<a href="#">37</a>
<a href="#">Figure 4.3 The Mixture Problem.....</a>	<a href="#">98</a>



## PREFACE

No graduate student is an island unto himself (or herself). I owe a great debt of gratitude to the many people who helped me as friends, mentors, advocates, patrons, and pranksters, through this rewarding, but sometimes trying, journey.

First, I must thank the institutions that supported my work. I was a member of Pitt's HPS Department and a graduate student member of the Center for the Neural Basis of Cognition. I also spent a year as a program assistant in the Center for Philosophy of Science, and received a fellowship from the Josephine De Karman Fellowship Trust. I am extremely grateful for their generous financial and professional support.

Thanks to my committee for all their help. Mazviita Chirimuuta gave me excellent, sharp philosophical criticism, and encouraged me to think more broadly about where fMRI stands in the economy of cognitive science. David Plaut brought his scientific expertise to the committee, as well as a great deal of philosophical acumen. David Danks joined late in the game, but quickly became a great mentor and collaborator who left a huge mark on both this document and on my development as a scholar. I would like to thank my advisors Edouard Machery and Peter Machamer for believing in me, for countless hours of guidance, and for always pushing me to think more deeply and insightfully about the topic at hand. And I have to say—Edouard, thank

you for taking me under your wing. If I accomplish anything in philosophy, it is because you took an interest in me, and patiently taught me the craft.

I also owe a great deal to Pittsburgh's HPS and neuroscience communities. Michael Tarr and Marlene Behrmann were kind enough to let me attend their VisCog laboratory meetings, which greatly enhanced my project. Thanks to Elissa Aminoff and John Pyles for great discussions (and to John for his fMRI course). On the HPS side, I benefitted from conversations with Jim Lennox, Ken Schaffner, Jim Woodward, Jim Bogen, Jeff Schwartz, Paolo Palmieri, John Norton, and Sandra Mitchell. And thanks to Joann McIntyre, Natalie Schweninger, Lisa Bopp, Rita Levine, and Joyce MacDonald for all of their support over the years: I couldn't have done this without them!

Thanks to my mentors beyond the Pitt community. A great talk by Michael Anderson spurred my interest in the project. Russ Poldrack, Tim Bayne, Nick Shea, Colin Klein, Jessey Wright, Mike Dacey, Bill Bechtel, Josh Alexander, Sarah Robins, Adrian Nestor, Carrie Figdor, and Tom Polger each gave me helpful advice and fostered my graduate research program. At Colorado College, Marc Snyder, Tass Kelso, Bob Lee, Lori Driscoll, Susan Ashley, Marion Hourdequin, Jonathan Lee, and Candace Galen, set me on the path to graduate school in philosophy and cognitive science.

My time in HPS was (despite its trials and tribulations) great. I owe this largely to the graduate students of HPS and Philosophy at Pitt, many of whom I count as true friends. Thanks to Tom Pashby, Katie Tabb, Julia Bursten, Elizabeth O'Neil, Eric Hatleback, Yoichi Ishida, Bihui Li, Greg Gandenberger, Lisa Lederer, Taku Iwatsuki, Marcus Adams, Lauren Ross, Nora Boyd, Aaron Novick, David Colaço, Morgan Thompson, Evan Pence, Alison Springle, Siska De Baerdemaeker, Zina Ward, Haixin Dang, Joshua Eisenthal, Mikio Akagi, and Trey Boone, for

everything. Thanks to Tom, Katie, and Julia for helping me through graduate school and the job market (and to Tom for “Beers and Gears”). Thanks to the current crop of HPS grads for cheering me up with all your zany antics (even if you did put a cockroach on my desk and sewed a cat to my jacket). And thanks to Trey and Mikio for being the friends I needed to get through these trials, and for all the nights of great philosophy, whisky, and camaraderie. Oh, and a quick shout-out to my friends beyond HPS. Thanks to Emre Ulkucu, Ian Bellayr, and Matt Miller for leaving Pittsburgh so I could finally get some work done. Thanks to Cameron Ritchey for putting up with weird phone calls (sorry about making you a human GPS buddy!), and to Alex Deverell, Mike Shum, Josh Redfearn, and Justin Patterson for being great friends and for having more faith in me than I do.

A special note of thanks goes to my family. Thanks to Bob and Sue Thibadeau, Rob Thibadeau, and Anne and Gordon Greene for welcoming me into the family with open arms. Thanks to my brothers Sean McCaffrey, Brian Kuzma, and Hao Chien, for always having my back. Thanks to my sister Sarah Kuzma, for heaps of support and advice over the years on matters ranging from the profound to the mundane. And a huge thanks to my wife, Mary Thibadeau, for her endless patience and support, and also for being the best partner a man could hope for. I love you, Mary; without you, I never would have swum with sea turtles, petted a penguin, acted out plays in my living room, or realized how lucky a person can feel each day.

Last, I take this opportunity to offer the deepest thank to my parents. My mother Marilyn J. Curtis has supported me through all of life’s challenges with unfathomable compassion, patience, and love. My father Robert J. McCaffrey inspired me to wonder about the drifting stars and other secrets buried within this beautiful thing we call “nature.” He started me on the road to this dissertation, and traveled beside me as far as he could. I dedicate this dissertation to my

mother Marilyn, and to the memory of my father Robert.

## 1.0 DISSERTATION INTRODUCTION

### 1.1 PROJECT DESCRIPTION

Cognitive neuroscience is a scientific discipline that seeks to understand the brain basis of human thought and behavior. By its very nature, cognitive neuroscience spans many fields of scientific inquiry, including linguistics, artificial intelligence, psychology, and neurobiology. A central philosophical issue in cognitive neuroscience concerns whether it is possible—and if so, how—to integrate these diverse fields (Craver 2007, Sullivan 2009). Some researchers have argued that cognitive neuroscience is a “troubled marriage” of psychology and brain research (Cooper and Shallice 2010). That is, there is controversy over whether and how studying the brain informs our scientific understanding of mental processes such as thoughts or emotions (Hatfield 2000, Uttal 2001).

One major area of cognitive neuroscience, *human brain mapping* or *cerebral cartography*, has two broad goals. The first is to localize cognitive functions—broadly construed to include affective and perceptual processes—onto structures of the brain. For example, one might want to know what part(s) of the brain are responsible for “disgust,” “face recognition,” “working memory,” etc. The second is to use these mappings—e.g., the dorsolateral prefrontal cortex is involved in working memory—to test cognitive theories. For example, Gauthier and Tarr (2000) used neuroimaging to study whether humans possess a system dedicated to

recognizing conspecific faces. Functional magnetic resonance imaging (fMRI), a neuroimaging technique that measures changes in blood oxygenation due to neural activity, is currently the dominant tool in cognitive neuroscience for both of these purposes (Poldrack 2008, Aguirre 2014). That is, neuroimagers use fMRI to map functions onto the brain and to test psychological theories. But despite its popularity, the use of fMRI for these purposes is fraught with controversy. While some neuroscientists think we can use neuroimaging to revolutionize our conception of the mind (Anderson 2014, 2015), others have denounced neuroimaging as a “new phrenology,” a circular experimental paradigm that can tell us little (maybe nothing) new about the mind (Van Orden and Paap 1997, Uttal 2001).

How successful is the project of using fMRI for brain mapping, and for testing cognitive hypotheses? In other words, how does fMRI inform our best understanding of the functions performed by different parts of the brain, and how do fMRI findings bear on our best cognitive theories? Second, how are these projects related—that is, how does the use of fMRI for brain mapping inform the use of fMRI for testing cognitive hypotheses? These are foundational, yet unresolved questions in the philosophy of neuroscience and psychology, though many philosophers (Bechtel 2008, Roskies 2009, Klein 2012, Rathkopf 2013,) and philosophically minded neuroscientists (Coltheart 2004, 2006, Anderson 2010, 2014, Poldrack 2006, 2010, Chatham and Badre 2015) have recently written about them.

There is a great deal of philosophical interest in how functional attribution works in the biomedical sciences (e.g., Cummins 1975, Craver 2001) and on the research strategy of functional localization (e.g., Bechtel and Richardson 1993). And philosophers have sometimes weighed in on whether fMRI is valuable for testing cognitive theories (e.g., Fodor 1999). But there is currently no general account in the philosophical literature on the relationship between

human brain mapping and the value of fMRI for testing psychological hypotheses. My aim in this dissertation is to remedy this situation by advancing a novel account of the relationship between human brain mapping and cognitive theorizing in fMRI research. Of course, this dissertation owes a debt to many established philosophers (Bechtel 2008, Lloyd 2000, Roskies 2009, Klein 2010b, Machery 2012, 2013) whose excellent work on neuroimaging has paved the way for a new generation of philosophers (Rathkopf 2013, Nathan and Del Pinal 2016, Povich 2015) to write on these issues.

Part of my project is *descriptive*. I aim to demonstrate what assumptions about the brain—e.g., if some region is active in two circumstances, it is “doing the same thing” in both cases—underlie the inferences neuroscientists draw from fMRI experiments. Neuroscientists are not always explicit about the logic of their experimental practices, and this can make it difficult to critically assess them. Part of my project is *normative*. I also aim to critique current neuroimaging practices, illustrating what theoretical inferences are, or are not, permitted by current experimental techniques. I consider this dissertation an instance of “philosophy of science in practice.” That is, I attempt to show that genuinely *philosophical* analyses—e.g., theories of what functions are (Cummins 1975), or analyses of the relationship between theory and evidence (Nathan and Del Pinal 2016)—can help cognitive neuroscientists understand better understand their work and its limitations. Ultimately, I hope that my analysis will be useful for philosophers interested in the sciences of the mind and brain, philosophers who use fMRI findings to assess philosophical theories (e.g., empirically-oriented moral psychologists), and for cognitive scientists interested in the theoretical foundations of fMRI.

## 1.2 HISTORICAL AND PHILOSOPHICAL BACKGROUND

The question of how functions map onto the brain is one of the perennial issues in cognitive neuroscience. This is as true now as it was at the inception of the field. The 19<sup>th</sup> Century witnessed an explosion of scientific interest in the localization of functions in the human brain. Franz Gall (1758-1828), Jean Pierre Flourens (1794-1867), Paul Broca (1824-1880), David Ferrier (1843-1928), and many others all developed theories of how mental functions map onto the human cerebrum. Some, like Gall, were *localizationists* who believed that mental faculties—e.g., a sense of language or mathematics—could be localized to specific brain convolutions or gyri. Others, like Flourens, were *equipotentialists* who believed that the cerebrum acted en masse to produce thought and behavior (Mundale 2002). A brief analysis of Gall’s phrenological system will illustrate many of the issues still live in brain mapping research today.<sup>1</sup>

At a basic level, cerebral cartography involves relating a set of mental faculties or cognitive functions to a set of neural structures such as cerebral gyri, Brodmann’s cytoarchitecturally-defined regions, or widely distributed brain networks. This process involves: (1) A way of individuating or taxonomizing *mental faculties*, and (2) Experimental techniques and *rules for mapping* those faculties onto the brain. Franz Gall’s phrenological system offered a surprisingly modern way of individuating mental faculties. Where previous thinkers had largely relied on introspection to identify a small number of coarse-grained faculties such as “imagination,” “the intellect,” or “the will,” Gall developed a taxonomy consisting of dozens more abilities such as “sense of locality,” “faculty for words, verbal memory” and “arithmetic,

---

<sup>1</sup> Mundale (2002) helpfully distinguishes between equipotentialism (the view that cerebral cortex exhibits virtually no functional specialization) with holism (the view that nearly all of the brain is involved in performing most cognitive functions).



counting, and time” (Poldrack 2010). Gall identified these faculties based on observations that children could have poetic talent, but not musical talent, or that patients with brain injuries could exhibit impaired mathematical intelligence, but spared verbal intelligence. Then, Gall mapped these faculties onto the brain by correlating personality and intelligence traits with convolutions of the skull. While Gall’s method of individuating faculties is thought to have some merit, Gall was wrong that skull bumps correspond to cognitive capacities in any interesting sense (Anderson 2014, Ch. 1).

While some 20<sup>th</sup> century neuroscientists (e.g., Lashley 1933) have championed the idea that the cortex acts as a whole, most neuroscientists believe that the cortex exhibits a high degree of functional specialization, such that cognitive functions can be localized to individual regions (Kanwisher et al. 1997) or networks of regions working in a concerted fashion (Mesulam 1990, McIntosh 2000). The early experiments of Paul Broca, David Ferrier, and others won the day over holist alternatives. And while some psychologists (e.g., Uttal 2001) deny that it is possible to carve the mind at its functional joints, most believe that neuropsychological studies (Shallice 1988) and behavioral studies (Petersen and Fiez 1993) can reveal the existence of distinct cognitive components. Thus, while some psychologists denigrate the use of fMRI for brain mapping as a “new phrenology” (more on this below), the enterprise as it was conceived in the 1990s does resemble phrenology in its basic respects. That is, neuroscientists adopted a method where they individuated the components of the mind via behavioral studies or neuropsychology experiments (a taxonomic project), and then used neuroimaging to map these components onto the brain (a cartographic project). In this context, the main questions of my dissertation are: (1) How do cognitive neuroscientists develop their taxonomies of psychological kinds, (2) How do

neuroscientists map these kinds onto the brain?, and (3) Can facts about the brain actually influence our understanding of those cognitive kinds?

### **1.3 CONTROVERSIES ABOUT FMRI**

Functional MRI involves placing participants in a strong magnetic field (generally about 2 Tesla) and measuring local changes in blood oxygenation—called the blood oxygen dependent or “BOLD” response—in a set of brain “voxels” (volumetric pixels) (Banich and Compton 2010, Aguirre 2014). Fluctuations in the BOLD signal are considered a proxy for the metabolic activity of neural populations, such that some BOLD signal changes reflect changes in neural activation patterns as opposed to artifacts of cerebral vasculature flow, head motion, scanning procedures, etc. (Logothetis 2003). Standard univariate practices involve a obtaining an anatomical image, obtaining functional data, submitting these data to a number of corrections (e.g., spatial smoothing or head motion corrections), fitting the hemodynamic response to a general linear model (GLM), and finally, generating a statistical parametric map (SPM) of which voxels exhibiting significant changes to a contrast of interest (Buxton 2002, Banich and Compton 2010). It is important to note that fMRI does not produce functional “images” of the brain; the colored blobs in fMRI studies reflect statistically significant BOLD changes produced by applying statistics (including correcting for multiple comparisons, etc.) to a complex model of the raw data (Klein 2010a).

In the early days of neuroimaging (Petersen and Fiez 1993, Posner and Raichle 1998), cognitive neuroscientists developed an agenda for localizing cognitive functions onto the brain and then using these localizations to test psychological theories. In the first step, researchers

employ subtractive neuroimaging, which utilizes both cognitive subtraction and BOLD signal subtraction prior to statistical comparisons (see Banich and Compton 2010 for a more detailed description) to localize cognitive functions to particular brain areas. Henson (2006) calls this form of function-to-structure inference “forward inference.” For instance, researchers might subtract the patterns of brain activation in a standard speech comprehension task from patterns of activation in a Phonological Loop Task (a speech comprehension task designed to recruit working memory) to isolate activity related to working memory. This purportedly works because speech comprehension is a “matched control” for a Phonological Loop Task—that is, the tasks putatively differ only in terms of whether working memory is recruited or not. This assumption is known as “pure insertion” in the philosophical and scientific literature (Van Orden and Paap 1997).

In the second step, researchers use these localizations to test hypotheses in cognitive psychology. For instance, subtraction studies suggest that working memory elicits activation in the dorsolateral prefrontal cortex. Using this knowledge, researchers might design experiments designed to probe whether a novel task (e.g., doing long division) also recruits working memory by testing whether that task also elicits activation in the dorsolateral prefrontal cortex. This pattern of structure-to-function inference, which Poldrack (2006) calls “reverse inference,” is a common practice in cognitive neuroscience.

Together, forward inference and reverse inference present what we might call the “Standard Picture” of cerebral cartography. According to this picture, fMRI is good for testing cognitive theories in virtue of being good for localizing functions onto the brain. In the first pass—i.e. forward inference—neuroscientists localize cognitive functions—e.g., theory of mind—onto particular brain regions—e.g., right temporoparietal junction and associated regions.

In the second pass, researchers ask how these regions of interest respond to other stimuli or tasks. For example, Gauthier et al. (2000, see also Gauthier and Tarr 2000) ask whether the fusiform face area (isolated by contrasting faces versus other sufficiently complex objects) responds *just* to faces, or also other categories for which participants are experts (e.g., birds and cars). Or, researchers might see if a novel task—e.g., a moral judgment—recruits some psychological function of interest—e.g., disgust—based on whether it recruits regions associated with that function. The latter is an example of reverse inference.

There is a great deal of controversy about whether these practices rest on firm theoretical foundations. According to some theorists (Van Orden and Paap 1997, Uttal 2001), fMRI is not even useful for localizing cognitive functions onto the brain. These criticisms often hinge on the notion that the assumption of pure insertion used in subtraction studies —i.e. that researchers can identify *tasks that differ solely in the recruitment of one cognitive process or function*—assumes a simplistic, modular view of cognition. For example, a researcher might suppose that the only difference between the word, “DOG,” and the non-word, “BLORT,” is that one involves semantic access. This assumption might be illegitimate, since the word “DOG” not only involves semantic access, but perhaps also mental imagery, emotional processes, etc. Van Orden and Paap (1997) argue that pure insertion is not only unjustified; in many cases, it is demonstrably false (see Friston et al. 1997).

Another common criticism is that since subtraction studies will always reveal some peak BOLD signal difference (these peaks are usually obtained after correcting for multiple statistical comparisons), neuroimagers can localize whatever function they imagine to be the difference

between two tasks onto that activation peak (Van Orden and Paap 1997, Poldrack 2010).<sup>2</sup> In a provocative (perhaps humbling) thought experiment, Poldrack imagines that if Franz Gall had possessed an fMRI machine, he could have used BOLD changes to find the neural basis of sagacity, poetic talent, murderousness, aptness to receive an education, or virtually any of his faculties (2010, 754).

Other theorists argue that fMRI is useful for localizing cognitive functions onto the brain, but not for testing psychological hypotheses (Fodor 1999, Coltheart 2004, 2006). For example, Coltheart (2004, 2006) argues that since psychological theories do not make predictions about patterns of brain activation, neuroimaging results cannot—in principle—support or contradict a cognitive theory. By contrast, many researchers (e.g., Henson 2005) argue that fMRI is an indispensable tool for testing cognitive hypotheses. Some neuroscientists go so far as to claim that fMRI can be used to discover new psychological constructs (Anderson 2014, Ch. 4) or to revise our existing taxonomies of psychological kinds (Lenartowicz et al. 2010). Lindquist et al. (2012), for example, argue that fMRI results favor a social constructionist account of emotions over basic emotion theory. These proposals hearken back to the Churchlandian idea that familiar psychological categories will be eliminated or radically revised as brain research progresses (Churchland 1981). Thus there are deep theoretical disputes about the value of fMRI both for brain mapping and for testing cognitive theories.<sup>3</sup>

---

<sup>2</sup> See Roskies (2010) for a defense of subtraction against these charges.

<sup>3</sup> There are also disputes about specific experimental practices, rather than the value of the enterprise as a whole. For example, Poldrack (2006) argues that reverse inference is problematic since brain areas are activated in many different circumstances; Machery (2013) defends reverse inference against this charge.

## 1.4 OUTLINE OF DISSERTATION

My dissertation advances a novel account of the relationship between brain mapping and cognitive theorizing in fMRI research. I argue that fMRI studies can, in principle, inform the development and assessment of cognitive theories. Their ability to do so, however, hinges on bridging assumptions specifying how cognitive functions map onto the brain (see also Roskies 2009, Coltheart 2013, Nathan and Del Pinal 2016, Chatham and Badre 2015). The problem is that the landscape of human brain mapping is rapidly changing due to new techniques (e.g., resting state fMRI) and theoretical approaches (e.g., network-oriented mapping). These shifts undermine the assumptions—e.g., “each brain region performs a unique function”—that traditionally make fMRI results speak to cognitive theories. Therefore, many attempts to use fMRI to test psychological theories ultimately fail. I conclude that fMRI research should focus its research on developing new bridging assumptions rather than exclusively focusing on testing existing cognitive theories. Furthermore, the bridging assumptions that bring fMRI in contact with psychology will need to be sufficiently “local”—i.e. specific to inferences in particular experimental paradigms, about particular brain systems.

In [Chapter 2.0](#), I investigate the problem of multi-functionality in human brain mapping. Traditionally, cognitive neuroscientists have assigned one function (e.g., working memory or face recognition) to each brain region. But recent studies suggest that brain regions often have many different functions—for example, the cerebellum is involved not just in motor control, but also reading and spatial cognition. Multi-functionality undermines the search for systematic (i.e. selective) structure-function mappings. Philosophers and neuroscientists disagree about how brain mapping should proceed in light of multi-functionality. For example, Cathy Price and Karl

Friston (2005) argue that each region does perform a unique function, but that neuroscientists need to develop a new taxonomy of cognitive functions to capture what these regions are doing. In contrast, Colin Klein (2012) argues that neuroscientists should abandon the idea that each region performs a single function in favor of context-sensitive mappings. I argue that we are unlikely to find a general strategy for mapping functions onto multi-functional brain regions. Drawing on causal role theories of function in the philosophical literature, I argue that there is no canonical structure-function relationship for multi-functional components in biological systems, including the human brain. This “Functional Heterogeneity Hypothesis” holds that the brain contains different kinds of multi-functional parts; thus there is no general solution to the puzzle presented by multi-functional brain areas.

In [Chapter 3.0](#), I critique the “cognitive ontology revision” movement in cognitive neuroimaging. Recently, several prominent neuroscientists such as Russ Poldrack (2010) and Michael Anderson (2010, 2014, 2015) have argued that neuroimaging evidence compels a revision of psychological kinds. Throughout history, psychologists have often revised their psychological categories—e.g., “memory” was split into procedural and declarative variants. Contemporary neuroscientists often hold that fMRI should play an important role in such revisions—i.e. the lumping, splitting, etc. of cognitive kinds such as “working memory” or “anger.” For example, some neuroscientists claim that fMRI evidence contradicts basic emotion theory, which holds that emotions fall into discrete categories such as disgust, anger, fear, happiness, etc. If fMRI evidence should be used for the assessment of psychological categories, this has profound implications for the autonomy of psychology from neuroscience. I identify two strategies for conducting fMRI-based cognitive ontology revision. The first asks whether brain data “matches” predictions of psychological theories—e.g., basic emotion theory allegedly

predicts that regions involved in emotions will be specific to one category, such as anger or fear. The second takes a set of existing constructs (e.g., working memory versus top-down attention), and asks whether they elicit similar or different brain activation patterns. I argue that the first approach relies on overly strong bridging assumptions about how cognitive functions map onto the brain, while the second relies on untested methodological assumptions.

In [Chapter 4.0](#)—co-authored with David Danks—I examine the prospects of using resting state fMRI to identify functional brain networks.<sup>4</sup> In resting state fMRI, researchers collect data about what brain areas “communicate” with one another (called “functional connectivity patterns”) while participants rest in the scanner. What began as an accidental observation of patterned activity in the resting brain has become a major area of research. Many neuroscientists argue that the large-scale brain networks active at rest correspond to known brain networks (e.g., motor, visual, or attention networks), and thus resting state fMRI can be used to identify candidate functional brain networks in a “bottom up” fashion. This means researchers can discover new functional networks without making prior assumptions about the psychological differences between tasks, which is a radical departure from the stimulus-response paradigm characterizing much of cognitive psychology and neuroscience. I argue that resting state analyses involve sampling from a “mixture distribution” of unknown elements. After all, people lie in the scanner for hours, during which they perform any number of cognitive processes such as planning their day, remembering a vacation, or singing songs in their heads. The small number of large-scale, intrinsic networks identified in resting state research may reflect these psychological processes being blended together (across time) in the analysis rather than a feature of the brain’s functional organization. This suggests a need for methodological checks—e.g., statistical

---

<sup>4</sup> We contributed equally to the co-authorship of Chapter 4.



“demixing” methods, or more participant self-reports—before we can say with confidence that resting state fMRI reveals candidate functional brain networks.

In [Chapter 5.0](#), I conclude by describing the results of my analysis, and highlighting areas for further research. My main conclusion is that the value of fMRI for testing cognitive hypotheses depends on the validity of bridging assumptions about how cognitive functions map onto the brain. However, theories of the brain’s functional topography are shifting due to new methodological and theoretical developments. Thus, neuroscientists will need to evaluate their bridging assumptions. The upshot of this discussion is that the value of fMRI for testing cognitive theories is deeply tied up in our theories of the brain’s functional topography; as our conception of the brain changes, so too will the ways in which fMRI data bears on questions about the nature of the mind.

## 2.0 THE BRAIN'S HETEROGENEOUS FUNCTIONAL LANDSCAPE

### 2.1 CHAPTER 2 INTRODUCTION

Many neuroscientists assume that the cortex is a mosaic of functionally specialized regions, each of which performs a distinct cognitive function (Kanwisher 2010). Indeed, this “modular” picture informs a great deal of theorizing in cognitive neuroscience, either implicitly or explicitly. Current research suggests, however, that brain regions are often multi-functional (Anderson 2010, 2014). That is, the same anatomical structures are often recruited for very different cognitive functions. To illustrate, consider the cerebellum. While the cerebellum is traditionally viewed as a motor area involved in balance and coordination, recent studies implicate the cerebellum in numerous cognitive functions such as reading, habit formation, and spatial reasoning (Raberger and Wimmer 2003, Schmamann and Caplan 2006). How should neuroscientists interpret these findings? Does the cerebellum perform one function in connection with these diverse abilities, or does it merely perform different functions at different times?

Apparent cases of multi-functionality are ubiquitous in the human neocortex (Anderson 2010, 2014). This finding raises significant challenges for structure-function mapping in cognitive neuroscience. Chiefly, multi-functionality confounds the search for *systematic* or selective mappings between specific brain regions—e.g., the fusiform face area—and particular cognitive functions—e.g., face recognition. Neuroscientists and philosophers disagree about how

structure-function mapping should proceed in light of multi-functionality. Cognitive neuroscientists Cathy Price and Karl Friston (2005) argue that brain regions perform many functions at one level of description, and a single, previously uncharacterized, function at another. Thus neuroscientists need to revise their *cognitive ontologies*—i.e. taxonomies of cognitive functions—to obtain systematic mappings. Philosopher Colin Klein (2012) draws a very different lesson from the same findings: since the functions of brain regions vary according to the functional networks in which they are embedded, neuroscientists should adopt *context-sensitive* mappings rather than revise their cognitive ontologies. In other words, neuroscientists should give up on systematic mappings as an appropriate basis for theorizing (or accept that systematic mappings are not attainable for individual brain regions).

In this chapter, drawing on *causal role* theories of function in the philosophical literature (Cummins 1975, Craver 2001), I contend that neither account will succeed as a general treatment of multi-functionality in cognitive neuroscience: Brain regions, like other biological components (e.g., genes, tissues, organs, etc.) are multi-functional in different ways—viz., it is plausible that the brain contains *different kinds* of multi-functional parts. Furthermore, different kinds of multi-functional parts call for different functional mapping strategies. Therefore, there is probably no “one size fits all” solution to the puzzle presented by multi-functional brain regions. I call this the “Functional Heterogeneity Hypothesis.” As I will show later in the dissertation (particularly Chapters [3](#) and [5](#)), both the existence of multi-functionality, and the fact that there is no obvious solution to it, has profound implications for what neuroscientists can infer from fMRI results. Thus multi-functionality at the level of regions forms the basis for rethinking what it means for neural functions to map onto brain structures.

First, I discuss the problem of multi-functionality in greater detail ([2.2](#)). Then I examine

different strategies for dealing with multi-functionality in the brain [\(2.3\)](#). Next I argue that the value of different mapping strategies (e.g., cognitive ontology revision versus context-sensitive mapping) depends on the *mechanistic organization* (Machamer, Darden and Craver 2000, Craver 2001) of the target system—i.e. on the kinds of multi-functional components one is dealing with. Furthermore, I argue that the brain areas, like genes and organs, are probably multi-functional in different ways. Components with *conserved roles* perform the same basic operation in different capacities while components with *variable roles* perform different functions at multiple levels of description [\(2.4\)](#). I conclude by highlighting some broader implications for human brain mapping, and for the use of functional neuroimaging to investigate human cognition [\(2.5\)](#).

## 2.2 THE PROBLEM OF MULTI-FUNCTIONALITY

Human brain mapping is one of the main goals of cognitive neuroscience. Some hypothesized (though debatable) brain mappings include the right temporoparietal junction (RTPJ) and theory of mind, (Saxe and Kanwisher 2003), the fusiform face area (FFA) and face recognition (Kanwisher et al. 1997), and Broca's area and speech production (Tettamanti and Wenniger 2006). Structure-function mapping would be simplest if the brain obeyed a strict one-to-one relationship between anatomical regions (e.g., the RTPJ) and cognitive functions (e.g., mindreading). Unfortunately, the brain violates this tidy arrangement in two ways. First, many cognitive functions—especially fairly complex ones such as speech comprehension—are presumably carried out by *networks* of regions working in concert rather than any individual region (McIntosh 2000). Second, the same anatomical regions (e.g., Broca's area) often have

many different cognitive functions (Poldrack 2006, Anderson 2010). In other words, human brain areas are often multi-functional.

Numerous studies report a high degree of multi-functionality in individual brain regions (Anderson 2010, 2014). Recent functional magnetic resonance imaging (fMRI) studies link the insula to many functions including gustation (taste), pain sensation, disgust, empathy, attention, and working memory (Menon and Uddin 2010).<sup>5</sup> Broca's area, classically associated with speech production, has recently been implicated in additional functions such as syntactic processing, action imitation, and semantics (Tettamanti and Wenniger 2006). In a striking meta-analysis of neuroimaging studies, cognitive scientist Michael Anderson (2010) reports that cortical areas are "re-deployed" *on average* across nine different cognitive domains such as vision, audition, memory, and numeric cognition. Thus multi-functionality may be a global feature of the brain's functional organization.<sup>6</sup>

In what sense are brain areas multi-functional and why is this a problem for human brain mapping? Since there are many different concepts of biological function (see Wouters 2005), it is important to clarify the relevant notion. According to Robert Cummins' influential causal role account (1975), functional analysis involves decomposing a capacity  $\psi$  of some containing system  $S$  into a set of constituent operations or roles  $\Phi$ s that collectively perform that capacity.<sup>7</sup> For example, the capacity of breathing ( $\psi$ ) performed by the human respiratory system ( $S$ ) consists of roles including inhalation ( $\Phi_1$ ), exhalation ( $\Phi_2$ ), gas exchange ( $\Phi_3$ ), and the control of breathing rhythm ( $\Phi_4$ ). In this framework, the *function* of a component  $X$  is the *role* that  $X$  plays

---

<sup>5</sup> While many of the studies I cite are fMRI studies, multi-functionality is not just an artifact of cognitive neuroimaging. For example, neuropsychology studies corroborate the insula's multi-functionality (see Ibañez, Gleichgerricht and Manes 2010).

<sup>6</sup> Anderson's case for massive redeployment needs further evaluation since the extent of multi-functionality depends on the size of the regions investigated and the chosen categories of cognitive domains.

<sup>7</sup> I use the terms "role" and "operation" interchangeably to denote a sub-capacity  $\Phi$  of some target capacity  $\psi$ .

in S's ability to  $\psi$ —e.g., the brainstem's (X) role in breathing is the maintenance of breathing rhythm ( $\Phi_4$ ).<sup>8</sup>

Causal role functions nicely capture the sense in which brain areas are multi-functional.<sup>9</sup> Cognitive neuroscientists typically study complex cognitive capacities by decomposing them into a set of component operations—e.g., cognitive models of reading involve component operations such as attention, eye movement control, word form recognition and semantic access. Then neuroscientists use subtractive neuroimaging and other techniques to map these component operations onto the brain (Petersen and Fiez 1993).<sup>10</sup> Thus structure-function mapping in cognitive neuroscience often involves specifying the role  $\Phi$  a region plays in some cognitive capacity  $\psi$ . For example, the visual word form area's role in reading is to process the shape of written characters (Dehaene et al. 2005). But just as many organs are recruited for multiple physiological capacities—e.g., the liver functions in fat digestion ( $\psi_1$ ) and blood sugar regulation ( $\psi_2$ )—many brain areas are recruited for multiple cognitive capacities. For example, the insula is involved in both empathy ( $\psi_1$ ) and disgust ( $\psi_2$ ) (Menon and Uddin 2010).

Multi-functionality raises interpretive challenges for cognitive neuroscience (Price and Friston 2005, Anderson 2010, Klein 2012, Rathkopf 2013), particularly since the functions a region participates in often seem quite distinct from one another. To demonstrate, consider Broca's area. Since the 19<sup>th</sup> century, Broca's has been associated with speech production—i.e. the motor aspects of speech. However, recent studies implicate Broca's area in a variety of non speech-related functions. For instance, Heiser and colleagues (2003) found that rTMS (a

---

<sup>8</sup> This assumes that roles can be mapped onto a particular component; this condition might fail in certain complex systems, where functions may be attributed to the whole (see Bechtel and Richardson 1993, Ch. 2).

<sup>9</sup> This does not mean that every functional attribution in cognitive neuroscience designates a causal role. For instance, Garson (2011) may be right that some functional hypotheses concern a region's developmental history.

<sup>10</sup> Of course, not *every* sub-capacity will map onto a single region. For example, eye movement control is thought to rely on a distributed network of cortical and subcortical structures (see Liversedge et al. 2000).

technique that transiently “deactivates” brain tissue using magnetic pulses) of Broca’s area rendered participants unable to move their fingers in sequence with a recorded hand while preserving the basic ability to move their fingers. Meanwhile, Tettamanti and colleagues (2009) report that Broca’s area is involved in visuospatial learning. What does the recruitment of Broca’s area for speech production, action imitation, and visuospatial learning suggest about its function? Does Broca’s area perform a single function explaining its involvement in these different capacities, or does it perform different functions in different contexts? Simply listing the various cognitive functions associated with a region cannot resolve these issues. Therefore, neuroscientists need new strategies for functional mapping.

### **2.3 FUNCTIONAL MAPPING STRATEGIES**

Human brain mapping often aims at “systematic” structure-function mappings (Price and Friston 2005). Systematic mappings are those in which each structure of interest (e.g., dorsolateral prefrontal cortex) is uniquely associated with a particular cognitive function (e.g., working memory). According to a traditional view of brain mapping, the cortex is a patchwork of regions, each performing a single, unique function. The research I reviewed above suggests that this picture is wrong: Each region (as we know them) seems to perform many different cognitive functions (as we know them). There are many ways of responding to this situation. Perhaps there are systematic structure-function mappings, but not at the level of individual brain regions. Or perhaps brain regions *do* each perform a unique function, but we have the wrong set of functions for capturing what they are doing. Now I examine the strategies neuroscientists might adopt for dealing with multi-functionality, focusing on a recent debate between Price and Friston (2005)

and Klein (2012). First, I briefly discuss a strategy that seeks to *explain away* the problem by dividing each multi-functional area into smaller functionally specific regions.

### **2.3.1 Subdivide and conquer**

One strategy for dealing with multi-functionality is to divide the region in question into multiple functionally specific areas (e.g., Grill-Spector, Sayres, and Ress 2006, Fedoranko, Duncan, and Kanwisher 2012). According to this “subdivide and conquer” strategy, brain regions only *appear* multi-functional since current ways of delineating brain regions (e.g., BOLD dissociations in fMRI or cytoarchitectural maps) lump functionally distinct neural populations together (Grill-Spector, Sayres, and Ress 2006). For example, while Mitchell (2008) claims that the RTPJ is involved in both attention and theory of mind (the ability to infer other agents’ mental states), Scholz and colleagues (2009) argue that different sub regions of the RTPJ perform these different functions. Similarly, Wager and Barrett (2004) argue that the insula contains distinct sub-regions involved in core affect (e.g., disgust expression), motivation and cognitive control, and pain.

New techniques such as fMRI adaptation (Price and Friston 2005) and high-resolution fMRI (Grill-Spector, Sayres, and Ress 2006) offer promising avenues for sub-dividing multi-functional regions. However, it is unlikely that this strategy will explain away every instance of neural multi-functionality. First, in practice, multi-functionality often resists subdivision—e.g., the insular sub regions identified by Wager and Barrett (2004) still perform multiple functions. Second, multi-functionality is widely considered a basic feature of neural organization observed both in well-characterized invertebrate circuits (see Getting 1989) and at small spatial scales in the primate brain (see Anderson 2014, Ch. 2). Therefore, it is plausible that the human brain



contains *genuinely* multi-functional components—i.e. multi-functional populations that resist further subdivision. Assuming that some regions are genuinely multi-functional, how should brain mapping proceed?

### **2.3.2 Cognitive ontology revision: new systematic mappings**

Cathy Price and Karl Friston (2005) offer an alternative strategy for analyzing multi-functional brain regions. They argue that brain regions perform many functions at one level of description and a single function at another. The basic idea is that while a region might participate in many different cognitive capacities—e.g., reading, mathematics, attention, etc.—each region also performs a *single function* (i.e. a characteristic computation or operation) that explains its recruitment across these different contexts. Neuroscientists typically describe the functions of brain areas using functions such as “working memory,” inherited from cognitive psychology. Price and Friston contend that these traditional psychological kinds cannot provide systematic mappings for multi-functional regions—i.e. a region will always be multi-functional when described in terms of familiar psychological processes. Instead, neuroscientists need to revise their cognitive ontologies—i.e. develop new taxonomies of cognitive functions—to obtain systematic functional mappings. In other words, each region has many context-sensitive functions and one context-invariant one. Capturing these context-invariant functions—what Charles Rathkopf (2013) calls “intrinsic functions”—will require developing new cognitive kinds.

For instance, the posterior lateral fusiform gyrus (PLF) or “visual word form area” is hypothesized to function in word form recognition, a stage of reading in which the visual system

encodes the form of written characters (Dehaene et al. 2005). However, neuroimaging studies implicate PLF in many non-reading tasks such as categorizing pictures as animals or artifacts, recognizing objects by touch, and pairing visual cues with appropriate gestures (see Price and Friston 2005).<sup>11</sup> Thus “word form recognition” cannot provide a systematic mapping for PLF—e.g., it does not *predict* that categorizing pictures as animals or artifacts will recruit PLF or *explain* what PLF is doing when recruited for touch-based object recognition.

According to Price and Friston, the common denominator between PLF’s diverse functions is that in each case, “a [characteristic] motor response (name or action) is retrieved from [appropriate] sensory cues” (2005, 267). Therefore, Price and Friston argue that PLF performs the single function of “sensorimotor integration” in each context. Price and Friston stress that sensorimotor integration is a new *kind* of cognitive function because it is not a component of reading, object categorization, or tactile object recognition hypothesized in cognitive psychology. They hold that sensorimotor integration is more useful than traditional functions because it explains PLF’s recruitment in a wider range of experimental conditions.

Price and Friston draw a general lesson from PLF, arguing that the lack of systematic mappings in cognitive neuroscience stems from neuroscientists’ tendency to describe brain functions in familiar psychological processes instead of searching for deeper similarities outside the boundaries of traditional cognitive categories. They claim, then, that in addition to a region’s many context-sensitive mappings (e.g., PLF is for word form recognition, tactile object recognition, cue-action pairing), there is a systematic or context-invariant mapping (e.g., sensorimotor integration) capturing the region’s function in every context. The key for Price and

---

<sup>11</sup> That a region is active in some task (e.g., as measured by fMRI) is not sufficient for concluding that the region has some particular function. I will not worry about this complication while discussing this example.

Friston is that neuroscientists need to revise their cognitive ontologies to obtain these systematic mappings.

### **2.3.3 Context-sensitive functional mappings**

Colin Klein (2012) agrees that traditional cognitive functions cannot provide systematic mappings for multi-functional brain regions. However, Klein offers a very different take on the problem: neuroscientists should abandon systematic functional mappings in favor of context-sensitive ones—i.e. brain mappings that only hold in certain contexts. According to Klein, brain areas are flexibly recruited for different *functional brain networks*. So a region's function depends on what other parts of the brain are doing—i.e. on its “neural context” (see also McIntosh 2000). Neuroscientists must reference these “neural contexts” when mapping functions onto a brain region. Furthermore, Klein argues that Price and Friston's attempt to save systematic mappings through cognitive ontology revision (i.e. by articulating novel, context-invariant functions) is bound to yield uninformative mappings.

Klein argues that “sensorimotor integration”—Price and Friston's proposal for a context invariant function of PLF—is woefully uninformative. Many brain areas, including the parietal reach region, frontal eye fields, medial temporal area (MT), etc. are involved in pairing sensory cues with particular motor responses. As Klein notes, linking specific sensory cues to appropriate motor responses may be “what nearly all of cortex does” (2012, 955). Thus it is trivially true that the function of PLF is “sensorimotor integration,” since this could describe the function of virtually any brain area.

Klein elaborates that systematic functional mappings are often less illuminating than context-sensitive ones. To illustrate, Klein draws an analogy between brain regions and diesel truck pistons. Certain diesel truck pistons perform one of two functions ( $F_1$  or  $F_2$ ) depending on the context of the system ( $C_1$  or  $C_2$ ). Under normal driving conditions ( $C_1$ ), the piston compresses a fuel-air mixture as it moves upwards—this springs the piston down, which powers the engine ( $F_1$ ). When the engine brake is engaged ( $C_2$ ), exhaust valves release the compressed air before it springs the piston—the engine now drags the piston, which slows the truck down ( $F_2$ ). Klein argues that while both context-sensitive functional mappings (i.e. the piston performs  $F_1$  in  $C_1$  and  $F_2$  in  $C_2$ ) are useful, the piston’s only systematic mapping is a vacuous one like, “the job of the piston is to either speed the truck or to slow it down” (2012, 955).

Klein’s general lesson is that successfully mapping functions onto a component often involves specifying the broader context of the system in which the component is embedded. Whether the piston speeds or slows the truck depends on what other parts of the system, such as the engine brake and exhaust valves, are doing. In this situation, researchers should not search for systematic functional mappings (i.e. the one function the part performs in every context), but instead try to identify the *relevant contexts* anchoring different context-sensitive mappings. Klein suspects that this is true for brain regions, arguing that whether a region  $R$  performs one function or another critically depends on the network of regions with which it interacts. That is,  $R$  performs  $F_1$  in  $C_1$ ,  $F_2$  in  $C_2$ , etc. where the relevant contexts are neural contexts, or the set of regions with which  $R$  is co-activated.<sup>12</sup> For instance, PLF may perform a visual word recognition function in one context and a tactile object recognition function in another. Klein is not alone in

---

<sup>12</sup> As Klein (2012) notes, “neural context” refers to the *functional network* in which a region participates. Since the same anatomical network can exhibit different functional connectivity patterns, “sets of regions” are just proxies for functional networks, which include both what regions are activated and how those regions interact with one another.

raising this theoretical possibility, as many neuroscientists, such as Mesulam (1990) and McIntosh (2000), have also proposed that brain areas perform different functions in different contexts. Thus Klein argues that adopting context-sensitive mappings is a more promising strategy for brain mapping than revising one's cognitive ontologies to obtain systematic functional mappings.

## **2.4 DIFFERENT PARTS, DIFFERENT STRATEGIES**

We have seen three alternatives for dealing with multi-functionality. The subdivide and conquer strategy advocates splitting multi-functional areas into smaller, functionally specific regions. Price and Friston advocate revising one's cognitive ontologies, developing new functions such as “sensorimotor integration,” to regain systematic structure-function mappings. Finally, Klein proposes abandoning systematic mappings in favor of context sensitive ones. Presuming that the brain contains genuinely multi-functional parts—i.e. that the subdivide and conquer strategy will not always work—which approach is more promising for cognitive neuroscience?

I defend the view that neither strategy is more promising, and that each is necessary for progress in brain mapping. That is, we are unlikely to find a general strategy for mapping functions onto multi-functional brain regions. First, I argue that facts about the mechanistic organization of biological systems determine the value of different functional mapping strategies ([2.4.1](#)). I doubt that Price and Friston or Klein would, in principle, disagree, but in practice both advocate a general strategy for brain mapping without examining the facts about the brain that determine the success or failure of their preferred strategy. Drawing on causal role theories of function in the philosophical literature (Cummins 1975, Craver 2001), I argue that the relevant

facts determining whether cognitive ontology revision or context-sensitive mapping is more useful concern whether the region in question performs the same basic role or operation  $\Phi$  in different cognitive capacities  $\psi_1, \psi_2, \psi_3$ , etc. [\(2.4.2-2.4.4\)](#). I claim that the brain, like other biological systems (e.g., organs), contains different kinds of multi-functional parts—parts with *conserved roles* perform the same role in different capacities while components with *variable roles* do not. Since different parts call for different mapping strategies, there is likely no general solution to the problem of multi-functionality.

### 2.4.1 Mechanistic organization and functional mappings

Facts about the mechanistic organization of biological systems (Machamer, Darden, and Craver 2000, Craver 2001)—i.e. their spatial, temporal, and constitutive organization—determine the value of different functional mapping strategies.<sup>13</sup> To illustrate, consider the subdivide and conquer strategy.

The human pancreas has many functions including producing both the hormone insulin and enzymes used in fat digestion. Physiologists have discovered that two distinct cell populations carry out these functions. Cells in the Islets of Langerhans perform the pancreas' endocrine functions (e.g., insulin production) while acinar cells perform its exocrine functions (e.g., digestive enzyme production) (Fox et al. 2001, Ch. 18). The pancreas' multi-functionality is partly explained by the fact that it is a composite of distinct cell types devoted to different

---

<sup>13</sup> In this essay, I construe “mechanistic organization” in a minimal sense. While Cummins-style functional attributions (see Cummins 1975) do not require knowing every mechanistic detail of how components perform sub-capacities, they reflect basic organizational features such as the *hierarchical* structure of capacities and sub-capacities or roles (see Craver 2001).

functions—this is a triumph of the subdivide and conquer strategy. The human liver also has many functions including glycogen storage, bile secretion, and blood toxin removal. But unlike the pancreas, the liver is a fairly homogeneous mass of cells called “hepatocytes.” It is not the case that some hepatocytes are devoted to filtering the blood, while others are devoted to synthesizing glycogen. Instead, the same cells are capable of performing these different functions at different times (Fox 2001, Ch. 18).

So the subdivide and conquer strategy is useful for mapping functions onto what we might call *composite components*—i.e. components consisting of different kinds of functional subcomponents. Some neuroscientists think that traditionally defined cortical areas—i.e. Brodmann areas—will turn out to be such composites. The history of neuroscience suggests that has sometimes been the case. However, there is no guarantee that the subdivide and conquer strategy will always succeed since not all multi-functional components are composites. Some parts, such as the bulk of human liver tissue, have a homogenous composition that resists functional subdivision. Thus the value of the subdivide and conquer strategy depends on how the system in question is organized. I claim that the same is true of other kinds of multi-functional parts.

#### **2.4.2 Roles, capacities, and kinds of multi-functional parts**

What facts about the brain determine whether cognitive ontology revision or context-sensitive mapping is the right response to neural multi-functionality? Price and Friston and Klein dispute whether multi-functional brain areas perform a single function at another *level* of description. Klein understands these levels in terms of *abstraction*, writing: “Perhaps brain regions only appear pluripotent because we have not specified their functions in suitably general terms. Make

it abstract enough, and we will find that brain regions only do one thing after all” (2012, 954). According to this reading, cognitive ontology revision involves describing brain functions in more general terms until multi-functional areas become uni-functional. Just as walking and flying are both examples of “locomotion,” word form recognition and clapping in response to a visual cue are both examples of “sensorimotor integration.” Klein worries that such mappings achieve generality only at the expense of specificity and explanatory power. Saying that the function of PLF is “sensorimotor integration” is like saying that the hypothalamus’ function is “keeping people alive”—it covers everything, but explains very little.

I am sympathetic to Klein’s worry—abstract functional mappings do risk gaining generality at a loss of explanatory power, and “sensorimotor integration” is a prime example of this concern. Neuroscientists should not simply describe cognitive functions more abstractly until multi-functionality goes away. That said, pace Klein, the relevant levels of functional description for Price and Friston (and for the debate at large) are not levels of abstraction at all, but instead levels in a *mechanistic hierarchy* (Craver 2001). Causal role approaches to function distinguish between two levels of description for a component  $X$ : the broader capacity  $\psi$  to which a component contributes and the role  $\Phi$  that  $X$  plays in that capacity. For example, the valves ( $X_1$ ) in human veins contribute to circulation ( $\psi$ ) *by* preventing the backflow of blood ( $\Phi_a$ ), while the heart ( $X_2$ ) contributes to circulation *by* pumping ( $\Phi_b$ ). Biological components are often embedded in a functional hierarchy in which their immediate workings have broader functional consequences. Picture a gene  $G$ , which encodes some protein  $P$ . If  $P$  is involved in the inflammatory immune response  $R$ , then we might rightly say that the function of  $G$  is to encode  $P$  (at the level of roles), or that  $G$  is involved in response  $R$  (at the level of capacities). Talking about “levels” in terms of abstraction glosses over this distinction, and threatens to make the



debate about brain functions merely a verbal dispute. In contrast, multi-functionality raises interesting questions about levels of functional organization in the brain.

Return to the case of PLF, a brain region putatively involved in reading, tactile object recognition, and performing gestures in response to cues. According to Klein's reading, "sensorimotor integration," is just a more abstract description of these functions. In fact, "sensorimotor description" is not an abstract description at all, but a (perhaps unfortunate) name for a common role  $\Phi$  that PLF performs in each capacity. The claim is that PLF performs some basic operation—i.e. a computation or local working—that that is recruited for both reading and touch.<sup>14</sup> This is not a question about how abstractly to describe functions, but of whether PLF *does the same thing*—i.e. performs the same internal operation—when recruited for reading and cue-action responses. Many biological components are multi-functional in the sense that the same structure is recruited for different capacities  $\psi$ s. For instance, the brainstem is involved in both breathing ( $\psi_1$ ) and blood pressure regulation ( $\psi_2$ ). Whenever a part is recruited for different capacities, the question arises whether it performs the *same role*  $\Phi$  in each capacity ([Figure 2.4](#)). By "role," I mean the internal operation that a part performs. In what follows, I argue that biological components, including brain areas, differ in precisely this respect—i.e. parts with *conserved roles* perform the same basic operation in different capacities while parts with *variable roles* do not. Then I illustrate how these facts about brain organization bear on the value of cognitive ontology revision versus context-sensitive mapping.

---

<sup>14</sup> See Bergeron (2007) and Anderson (2010) for similar distinctions.

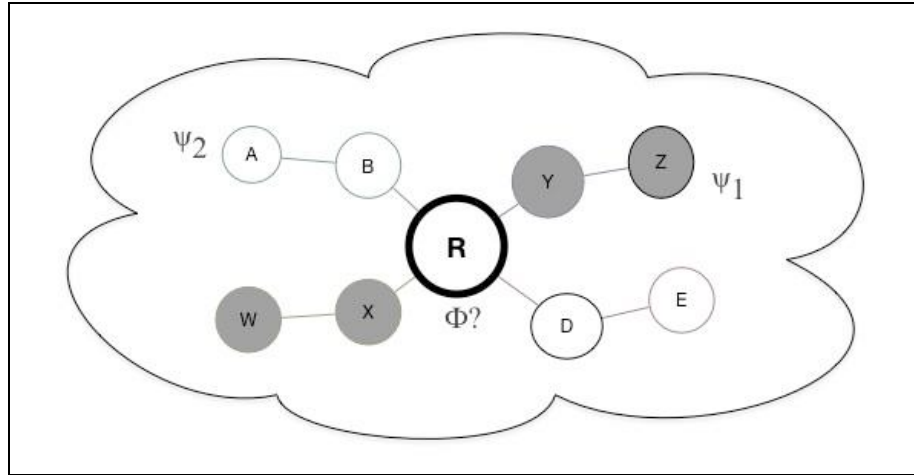


Figure 2.4: Roles versus capacities in brain mapping. The same brain region (R) can be recruited for functional networks for different capacities. For example, network  $\{W, X, R, Y, Z\}$  is associated with  $\psi_1$  while network  $\{A, B, R, D, E\}$  is associated with  $\psi_2$ . Region R is recruited for both networks, the question is: does it perform the same role  $\Phi$  in each capacity?

### 2.4.3 Conserved roles and systematic mappings

Some biological components have conserved roles—i.e. they perform the same basic role or operation  $\Phi$  in different capacities  $\psi_1, \psi_2, \psi_3$ , etc.<sup>15</sup> Consider the example of central pattern generators in systems neuroscience (Getting 1989, Briggman and Kristan 2008). In some species of leech, a single pool of motor neurons function as a central pattern generator that controls the

---

<sup>15</sup> I do not mean “conserved” in an evolutionary sense.

rhythm ( $\Phi$ ) of swimming ( $\psi_1$ ) and crawling ( $\psi_2$ ) motions (Briggman and Kristan 2008). Neurophysiological evidence suggests that while swimming and crawling rely on distinct muscle groups in the leech body, in both cases the pattern generator determines the rate and rhythm of the motion. But note that “rhythm control” is not a more abstract description of crawling and swimming—instead it denotes the common role that the central pattern generator plays in each capacity. Conserved role multi-functionality is found in other biological systems. Some genes exhibit “parsimonious pleiotropy,” a pattern in which “one gene is used for identical chemical purposes in multiple pathways” (Hodgkin 1998, 502).<sup>16</sup> In *E. coli* bacteria, a gene called *ilvN* encodes an enzyme (AHAS1) that is used in different pathways to synthesize either the amino acid valine, or the amino acid isoleucine (Dailey and Cronan 1986). In this case, a gene “does the same thing” for two different biosynthetic pathways. Thus many biological components perform the same basic operation in different capacities.

What makes a role “conserved” across different capacities will vary from system to system. In a computational setting, a string of code might execute the same subroutine ( $\Phi$ ) in different programs ( $\psi_1$ ,  $\psi_2$ , etc.). Klein (2012) mentions, for example, that a computational subroutine that performs fast Fourier transforms can be deployed for either an image compression program or an audio program. In a cell, an enzyme might catalyze ( $\Phi$ ) a reaction shared by different metabolic pathways ( $\psi_1$ ,  $\psi_2$ , etc.). Not all role attributions are particularly useful. A brain region  $R$  might perform the same role of “secreting neurotransmitters” in different cognitive capacities, but this functional characterization is not very informative. Attributing a conserved role to some component  $X$  will involve demonstrating that the component performs a sub-capacity ( $\Phi$ ) does explanatory work in analyzing a target capacity

---

<sup>16</sup> Pleiotropy occurs when one gene affects multiple phenotypes.

( $\psi_1$ ), and then demonstrating that this same role ( $\Phi$ ) does explanatory work in analyzing a another capacity ( $\psi_2$ ) with which the component is associated. For example, once we understand that the leech central pattern generator controls the pace and rhythm of different motions, we can predict that certain interventions on the pattern generator, such as a pharmacological blocker, will slow the rhythm of both crawling and swimming. The norms governing informative role attributions (as opposed to trivial or gerrymandered ones) will vary in different scientific fields.<sup>17</sup>

Components with conserved roles permit systematic mappings of the kind Price and Friston envision—that is, at the level of roles, the component does perform just one function. This formulation holds that brain areas *typically* perform the same basic role or operation  $\Phi$  in different cognitive capacities  $\psi_1$ ,  $\psi_2$ , etc. Other authors have proposed similar theories about multi-functional regions. For example, Anderson (2010, see also Bergeron 2007) distinguishes between the diverse higher cognitive functions for which regions are recruited (e.g., reading) and the re-deployed “workings” or computations the regions perform. Do brain areas typically perform the same role in different cognitive capacities?

Recent work suggests that the intraparietal sulcus (IPS), a region implicated in numeric cognition in humans and other mammals (e.g., rats and monkeys) performs the same basic operation in judgments of number and time. In humans, neuroimaging results suggest that IPS is recruited for judgments of quantity (e.g., does pile A or pile B contain more dots?) duration (e.g., did tone A or tone B sound for longer?), and size. A leading theory of these effects is that IPS implements a common mechanism for representing analog magnitude ( $\Phi$ ) that is flexibly

---

<sup>17</sup> Some philosophers (e.g., Neander 1991) worry that role individuation in causal role accounts is hopelessly unconstrained, hence rendering the approach trivial. Following Amundson and Lauder (1994), I hold that scientific practice provides principled grounds for role individuation in particular domains of inquiry. But I will not defend the point in detail here.

recruited for visual estimations of quantity ( $\psi_1$ ) and auditory estimations of duration ( $\psi_2$ ) (Pinel et al. 2004, Bueti and Walsh 2009).

Many other brain regions are hypothesized to perform the same role in different cognitive capacities. For example, human dorsolateral prefrontal cortex is thought to implement a selection mechanism ( $\Phi$ )—i.e. a mechanism for choosing which stimuli will be rehearsed—for two distinct working memory networks: the visuospatial sketchpad and the phonological loop ( $\psi_2$ ) (Baddely 2003). Aminoff, Kverga, and Bar (2013) argue that the parahippocampal cortex (PHC) plays the same basic role in several different cognitive capacities. The PHC is involved in many functions including spatial memory, visual scene processing, and even non-spatial forms of episodic memory (e.g., odor-odor associations in the rat). Aminoff and colleagues propose that PHC performs a form of contextual processing ( $\Phi$ )—i.e. accessing associative links in long-term memory—for different capacities such as scene processing ( $\psi_1$ ) and non-spatial episodic memory ( $\psi_2$ ). Thus conserved role multi-functionality is found in many biological systems and is hypothesized to exist in the human brain. Where these roles have not been discovered, characterizing them may call for developing new kinds of cognitive functions. Just as IPS appears to make the same contribution—i.e. analog magnitude representation—to duration and quantity judgments, other multi-functional areas—e.g., PLF—may have conserved roles that we have not yet characterized.

#### **2.4.4 Variable roles and context-sensitive mappings**

As we have seen, some biological systems have parts that play the same basic role in different capacities. Discovering these roles may produce informative, systematic mappings of the kind

Price and Friston envision. But just as the subdivide and conquer strategy will sometimes fail because not all multi-functional parts are composites, cognitive ontology revision may fail if there are components without conserved roles. Do brain regions perform the same operation, regardless of what psychological capacity they are recruited for? Should the goal of cognitive neuroscience be to discover each area's context-invariant function?

If Klein's analysis of diesel truck pistons is right, then some components exhibit context-sensitivity at both the level of roles  $\Phi$  and capacities  $\psi$ s. Under normal conditions ( $C_1$ ), the piston speeds the truck up ( $\psi_1$ ) by compressing a fuel-air mixture ( $\Phi$ ); when the engine brake is engaged ( $C_2$ ), the piston slows the truck ( $\psi_2$ ) by acting as a weight that the engine drags ( $\Phi$ ). Klein's analysis here points to an interesting feature of biological organization: parts can be multi-functional at different levels in a functional hierarchy. Some multi-functional components have *variable roles*—i.e. they perform different roles or operations  $\Phi_a$ ,  $\Phi_b$ ,  $\Phi_c$ , etc. in different capacities  $\psi_1$ ,  $\psi_2$ ,  $\psi_3$ , etc.

Variable role multi-functionality is found in many biological systems. Some regulatory genes can enhance or repress transcription depending on their biochemical context (Hodgkin 1998). For example, the gene *Ultrabithorax* modulates leg segment growth in some aquatic insects (e.g., water striders). However, due to differential regulatory effects, *Ultrabithorax* expression shortens some developing leg segments while lengthening other developing segments (Khila, Abouheif, and Rowe 2014). Similarly, liver hepatocytes perform different roles in different capacities. When the pancreas produces insulin, hepatocytes absorb ( $\Phi_a$ ) glucose from the bloodstream for glycogen synthesis, which lowers blood sugar ( $\psi_1$ ). At other times, hepatocytes secrete bile ( $\Phi_b$ ), which contributes to fat digestion ( $\psi_2$ ). Thus some components contribute to different capacities in virtue of different operational capabilities.

Components with variable roles often lack systematic mappings. Liver cells do not “do the same thing” when they contribute to fat digestion and blood sugar regulation, except in a very general or abstract sense (i.e. metabolism). Context-sensitive mappings are often more useful for such parts. This interpretation holds that brain areas *typically* (or at least enough to trouble the cognitive ontology revision approach) perform different roles  $\Phi_a$ ,  $\Phi_b$ , etc. in different cognitive capacities  $\psi_1$ ,  $\psi_2$ , etc. Is there evidence of variable role multi-functionality in the brain?

Recent work in cognitive neuroscience suggests that the same neural populations can implement different coding schemes for different channels of environmental information. The hippocampus is involved in both spatial navigation ( $\psi_1$ ) and episodic memory ( $\psi_2$ ). Leutgeb and colleagues (2005) argue that a single population of hippocampal neurons (found in CA1-CA3) has distinct signaling patterns for these capacities. According to their model, what subset of the population is firing ( $\Phi_a$ ) signifies the rat’s spatial location ( $\psi_1$ ); these “place field configurations” correspond to location, but not environmental features (e.g., colors, shapes, etc.). At the same time, the population’s rate function ( $\Phi_b$ ) reflects the presence of certain environmental features ( $\psi_2$ ) regardless of the active place configuration. Therefore, these neurons implement population coding for spatial memory and rate coding for episodic memory.

Other regions are thought to perform different roles in different cognitive capacities—e.g., while the dorsal striatum is involved in reward learning ( $\psi_1$ ) and voluntary movement ( $\psi_2$ ), models of its contribution to these capacities involve distinct operations. The dorsal striatum’s role in reward learning involves temporal difference detection ( $\Phi_a$ )—i.e. detecting differences between anticipated and actual reward onset—while its role in initiating movements is modeled as a simple disinhibition or gating mechanism ( $\Phi_b$ ) (Suri and Schultz 2001, Liljeholm and O’Doherty 2012). Likewise, primate neurophysiology studies suggest that certain inferotemporal

cortex (IT) neurons exhibit one spiking pattern ( $\Phi_a$ ) for detecting global features of objects ( $\psi_1$ ) and another spiking pattern ( $\Phi_b$ ) corresponding to local features ( $\psi_2$ ) of the same stimuli (Wang, Tanifuji, and Tanaka 1998). These results suggest that some brain areas perform different roles in different cognitive capacities. Unlike regions such as the PHC or IPS that seem to perform a common operation in each capacity, these areas are multi-functional at multiple levels of description. Thus there is no guarantee that cognitive ontology revision, construed as the search for systematic mappings at the level of roles, will succeed.

## **2.5 THE FUNCTIONAL HETEROGENEITY HYPOTHESIS**

There is no canonical structure-function relationship for multi-functional components in biology. Given recent research in cognitive neuroscience, this appears to be equally true of genes, organs, and neural systems. This suggests a “Functional Heterogeneity Hypothesis,” which holds that the brain contains different kinds of multi-functional parts. According to this hypothesis, the brain exhibits a heterogeneous functional organization in which different regions are multi-functional in different ways.

The value of different mapping strategies in biology is inextricably tied to the mechanistic organization of the target system ([see Figure 2.5](#)). For instance, composite components (e.g., the pancreas) are amenable to the subdivide and conquer strategy while other multi-functional components are not. The Functional Heterogeneity Hypothesis predicts that there will be no general account of how brain mapping will proceed in light of multi-functionality. Where regions have conserved roles, characterizing these roles (e.g., analog magnitude representation or contextual processing) may yield novel systematic mappings



reflecting interesting similarities between seemingly disparate cognitive capacities. However, there is no guarantee that this strategy will work. To the extent that brain regions have variable roles (e.g., implementing different coding schemes for different stimuli), context-sensitive mappings will often prove more useful. Which strategy is preferable will depend on the target brain region.

Type of component	Mapping Strategy	Examples
Composite	Subdivide and Conquer	Pancreatic tissue, Insula
Conserved Role Multi-Functional	Systematic Mapping (via functional revision)	Leech Central Pattern Generator, Intraparietal Sulcus
Variable Role Multi-Functional	Context-Sensitive Mapping	Liver tissue, Hippocampus, <i>Ultrabithorax</i>

Figure 2.5: The Functional Heterogeneity Hypothesis. Different *mapping strategies* are useful for mapping functions onto different *kinds* of multi-functional components.

If my account is right, then the challenge for neuroscientists is not to provide a general strategy for mapping functions onto multi-functional regions, but strategies that work for particular kinds of areas and ways of identifying what kinds of areas there are. Depending on factors such as the size of the region in question, one's existing taxonomy of cognitive functions, and the mechanistic organization in of brain systems, progress in brain mapping might require cognitive

ontology revision, new methods of subdividing regions, or context-sensitive mapping. The trick is determining when each approach is needed.

### 3.0 COGNITIVE ONTOLOGY REVISION AND fMRI: TWO STRATEGIES

#### 3.1 CHAPTER 3 INTRODUCTION

Psychologists typically study human cognition indirectly by positing the existence of hypothetical mental faculties or psychological constructs.<sup>18</sup> Following Price and Friston (2005), I will refer to the taxonomy or set of psychological constructs taken to be legitimate components of the mind as a “cognitive ontology.” Sigmund Freud’s (1859-1939) cognitive ontology included the id, ego, and superego. Franz Gall’s (1758-1828) cognitive ontology included a faculty of language, mechanical skill, poetic talent, and metaphysical perspicuity. Cognitive ontologies today include functional components such as spatial cognition, theory of mind, episodic memory, and so forth (Poldrack 2010).<sup>19</sup>

Historically, there has been a great deal of controversy about: (1) *What* the right cognitive ontology is (2) *How* scientists can determine what the right cognitive ontology is—viz., what set of experimental, inferential, and statistical methods provide evidence about our cognitive

---

<sup>18</sup> As I discussed in Chapter 2, these construct labels—e.g., “working memory,” “top-down attention,” etc. can often be thought of as verbal proxies for particular mechanistically or computationally defined capacities.

<sup>19</sup> This use of the term “ontology” here derives from informatics, where it designates a taxonomy of relationships between concepts or kind terms, rather than from philosophy (see Price and Friston 2005).

ontologies.<sup>20</sup> Note that “cognitive ontologies” is just a fancy way of talking about the taxonomy of psychological kinds; debates about cognitive ontologies are essentially debates about what philosophers would call the “natural kinds” of cognitive psychology (see Khalidi 2013, Polger and Shapiro 2016). In contemporary terms, there is disagreement about both the *validity* of particular psychological constructs or kind terms such as “anger” or “response inhibition” (Lenartowicz et al. 2010, Lindquist et al. 2012) and about the *methods* used to validate psychological constructs.

Recently, a number of prominent cognitive neuroscientists have argued that fMRI studies necessitate cognitive ontology revision (Price and Friston 2005, Poldrack 2010, Lenartowicz et al. 2010, Anderson 2014 Ch. 4, 2015). The basic idea is that neuroimaging findings should compel psychologists to revise their psychological kinds. In other words, fMRI bears on the validity of psychological constructs such as working memory or disgust. For example, Lindquist and colleagues (2012) argue that fMRI findings pose a significant problem for basic emotion theory, which divides emotions into discrete categories such as anger, happiness, sadness, surprise, and disgust. Similarly, Lenartowicz and colleagues (2010) argue that fMRI results validate some cognitive control constructs and cast doubts on others.

A couple notes about cognitive ontology revision are in order. First, the debate is not just about verbal labels such as “anger” or “working memory,” but about the mechanistic specifications of such labels (i.e. about how to individuate cognitive processes). For example, some neuroscientists hold that “response inhibition, “which involves suppressing a prepotent response,” is a different process than “task-switching,” which occurs when participants move

---

<sup>20</sup> Of course, this way of formulating the problem presumes that there is a single, correct taxonomy of psychological kinds. As Danks (2015), and others, have argued, pragmatic considerations might necessitate carving up our scientific ontologies differently for different purposes.

from one template to another. By contrast, Lenartowicz et al. (2010) use fMRI to claim that these verbal labels may denote the same brain process. Second, the taxonomy of psychology does not consist only of numerically distinct cognitive processes. For example, psychologists talk about “memory” as a set of cognitive processes including (at the very least) declarative memory, working memory, episodic memory, recognition memory, procedural memory, etc. Thus some of the kind terms in psychology refer not to distinguishable cognitive processes, but to conceptually useful categories for grouping these processes—e.g., kinds like “memory,” “perception,” or “cognitive control,” each of which consist of many different cognitive processes. For the sake of this chapter, I will refer to distinguishable cognitive processes as the “constructs” of one’s cognitive ontology.

In its most extreme form, harkening back to the Churchlandian idea that psychological categories will be eliminated as neuroscience progresses (Churchland 1981), researchers envision a scenario where neuroimaging spurs a revolution in conceptualizing psychological kinds (Anderson 2015). In its modest form, researchers consider fMRI a fruitful source of appeal for competing cognitive theories (Chatham and Badre 2015). If fMRI can be used to assess construct validity in psychology, this has profound philosophical implications. If fMRI can be used to assess construct validity in psychology, this has profound philosophical implications. For one, theories pitched at the so-called “psychological” level would not be immune to findings about neural implementation, as some theorists have suggested (Fodor 1999, Coltheart 2004, 2006).<sup>21</sup> Second, this would reflect a major methodological shift in cognitive science in which a new form of brain data could contribute to the development and assessment of psychological constructs.

---

<sup>21</sup> Of course, even if fMRI cannot assess psychological construct validity, other forms of neuroscientific evidence may prove useful for evaluating psychological constructs.

In this chapter, I investigate whether fMRI is a legitimate tool for assessing the validity of psychological constructs. I examine two strategies for cognitive ontology revision. The first strategy (Greene et al. 2001, Lindquist et al. 2012) tests one’s cognitive ontology by assuming a perspicuous mapping between psychological constructs, on one hand, and parts of the brain on another—e.g., that unique psychological functions are carried out by unique parts of the brain. In this strategy, neuroimagers implicitly or explicitly use *bridging assumptions* (see Roskies 2009, Coltheart 2013, Nathan and Del Pinal 2016 for a discussion) specifying how cognitive functions map onto neural structures to test the validity of psychological constructs.<sup>22</sup> The problem with this approach is that the bridging assumptions that neuroscientists use to link fMRI data to cognitive theories are often implausible and overly strong. The second strategy uses pattern classifier techniques such as multi-voxel pattern analysis (MVPA) to draw inferences about whether tasks recruit similar or distinct cognitive processes (e.g., Lenartowicz et al. 2010). While this strategy typically relies on weaker bridging assumptions about structure-function mapping, it involves problematic methodological assumptions about what neuroscientists can infer from classifier performance. I conclude that fMRI studies provide very weak evidence about the validity of psychological constructs, even if it is useful for other purposes (e.g., testing the relationship between constructs).

Here is how I will proceed. First, I clarify some of the main conceptual issues surrounding cognitive ontology revision (3.2). Here, I place cognitive ontology revision in the context of construct validation and show some of the different ways that cognitive ontology revision might proceed. Second, I discuss the case for fMRI-based cognitive ontology revision,

---

<sup>22</sup> Nathan and Del Pinal (2016) use the term “bridge laws” in their essay. I prefer “bridging assumptions,” since neuroscientists do not take the generalizations linking cognitive processing to fMRI findings to be necessary, universally true, etc. I think these bridging statements operate more like heuristics or assumptions guiding research (Bechtel and Richardson 1993).

highlighting different ways that neuroscientists have used fMRI to draw inferences about psychological kinds. Here I highlight that fMRI results can, in principle, bear on construct validity, but their ability to do so depends on auxiliary hypotheses or bridging assumptions about the brain (3.3). Then, I argue that the bridging assumptions that neuroimagers use to test cognitive ontologies are problematic given recent findings in human brain mapping (3.4). Finally, I introduce and critique pattern classification as a means of resolving some of these issues (3.5). I conclude with some remarks on the use of fMRI for cognitive ontology revision (3.6).

### 3.2 COGNITIVE ONTOLOGY REVISION

Psychologists typically study cognition indirectly through *operationalized* measures. For example, the Stroop Task is taken to recruit cognitive control (Sabb et al. 2008). The general problem of *construct validity* concerns how researchers can make correct inferences from operationalized measures to the nature of the *constructs*—i.e. unobservable mental processes or entities—taken to underlie those measures (Cronbach and Meehl 1955, Cambell and Fiske 1959, Smith 2005). A psychologist might want to know if the Stroop Task recruits the same cognitive control process(es) as the Go/NoGo Task. Psychologists assess construct validity using a variety of methods for comparing operationalized measures to theoretical predictions about the relationship between constructs. For instance, establishing *convergent validity* involves demonstrating that two independent measures of one putative construct (e.g., two different cognitive control tasks) are related to one another while establishing *discriminant validity* involves showing that independent measures associated with putatively different constructs (e.g.,

a cognitive control task versus a memory task) are unrelated (Campbell and Fiske 1959, Campbell 1960).

I will not delve deeply into the statistical issues surrounding construct validation. Instead, I am pointing out that the debate about cognitive ontology revision is essentially a debate about how experimentally observed, task-related changes in the BOLD signal (i.e. what fMRI measures) relate to unobservable psychological kinds or constructs (again looking at instances where these constructs refer to distinct psychological processes rather than “higher level” taxonomic groupings such as “emotion” or “perception”). In a broad sense, the project of brain-based cognitive ontology revision requires that there is some way to relate the neural events measured by the BOLD signal to taxonomies of cognitive kinds. Bracketing the issue of neuroscience for the moment, most psychologists—even behaviorists, insofar as learning rules are psychological kinds—implicitly or explicitly endorse some particular “cognitive ontology,” or set of psychological constructs taken to be valid psychological kinds.<sup>23</sup> There is often significant disagreement about how to carve the mind at its functional joints; for example, basic emotion theorists (e.g., Ekman 1999) hold that “anger” is a valid psychological construct, while social constructionists (e.g., Lindquist et al. 2012) do not. For example, emotional constructivists hold that different instances of what we call colloquially call “anger” do not involve the same set of “psychological ingredients” (see Lindquist et al. 2012). These disputes are about what *members to include in one’s cognitive ontology* often lead psychologists to revise their taxonomies of psychological kinds.

---

<sup>23</sup> It is possible that some psychologists adopt an anti-realist stance toward cognitive constructs such as “attention” or “working memory.” Another complexity is that there is likely substantial taxonomic variation between individual psychologists and different psychological research traditions.



There are many ways in which cognitive ontology revision can proceed. In divergence or *kind splitting* psychologists adjust their ontologies so that one kind—e.g., “memory”—is split into two—e.g., procedural and declarative memory (Craver 2004, Polger and Shapiro 2016). In convergence or *lumping*, psychologists revise their ontologies so that different kinds are lumped together. For instance, Gauthier and Tarr (2000) argue that “face recognition” is not a distinct psychological capacity, but one manifestation of a more general capacity for visual expertise (i.e. discriminating between different exemplars of a visual category).<sup>24</sup> Other times, psychologists revise their ontologies through *discovery*—i.e. adding new cognitive kinds. For example, Baddeley (2000) added a component called the “episodic buffer” to an initial model of working memory (Baddeley and Hitch 1974) in order to account for incongruous findings such as the phonological similarity and word length effects. Finally, some psychological kinds, such as Gall’s “metaphysical perspicuity” or Freud’s “death impulse” are *eliminated* after further research.<sup>25</sup>

Thus cognitive ontology revision occurs when psychologists revise their taxonomies by lumping and splitting cognitive kinds, by adding new members via discovery, or by eliminating existing kinds.<sup>26</sup> Now, I need to make a couple of quick remarks about the previous discussion of cognitive ontology revision in [Chapter 2](#). First, reassessing the functions of brain areas—e.g., that the superior temporal sulcus detects social situations rather than “biological motion”

---

<sup>24</sup> Another way to interpret Gauthier and Tarr’s (2000) claim is that face recognition (after learning) is a distinct psychological capacity from other categories, but one that relies on a general learning system.

<sup>25</sup> Paul Thagard (1990, 2012) adopts a similar framework outlining the ways in which conceptual change occurs in science.

<sup>26</sup> This is an overly simplistic picture, as there is room for reconceptualizing cognitive kinds (e.g., achieving a different understanding of what “working memory” is) without wholesale changes to one’s cognitive ontology.

(Pelphrey et al. 2004)—need not always amount to cognitive ontology revision.<sup>27</sup> Establishing whether the right temporoparietal junction is involved in top-down attention or theory of mind would not (for most researchers) result in cognitive ontology revision, since these are standard members of most cognitive ontologies. The claim that Price and Friston (2005), Anderson (2014, 2015) and others are making is that in some cases, reconceptualizing the functions of brain areas (particularly in ways that afford a mechanistic or computational understanding of how those regions operate) can lead to the discovery of new cognitive kinds. This is a special case of using fMRI to test one's cognitive ontologies; as we will see in the following section, there are other ways in which fMRI might contribute to cognitive ontology revision. Second, cognitive ontology revision need not involve brain data at all, since behavioral studies alone often support the revision of psychological kinds (Baddeley 2000).

In addition to debates about the validity of *particular psychological constructs*, there are issues surrounding the *methods* (broadly construed to include different instruments, experimental protocols, and different inferential or statistical interpretations of said results) used to validate and invalidate psychological kinds. Meehl (1978) portrays construct validation as a dynamic process in which psychologists articulate relationships between theoretical constructs, operationalize these constructs using various measures, compare predictions about these measures using different statistical techniques, and revise their taxonomies accordingly. Some debates about the methods used to validate or invalidate psychological constructs are debates about the *relevance* of particular instruments (and the dependent variables they measure) for assessing construct validity. Researchers have questioned, for example, whether the BOLD signal is really a good proxy of task-related neural activity. If the BOLD signal doesn't reflect

---

<sup>27</sup> I am not endorsing this interpretation of the superior temporal sulcus; I am just using it as a proposed example of functional revision.

neural activity, then presumably BOLD changes cannot bear on cognitive ontologies even if psychological processes are type-identical to neural ones (e.g., Logothetis 2008).

Other debates concern the logic of the experiments in question—viz., on the statistical and inductive inferences permitted by the findings. Here the question is not *whether* some dependent variable relates to hypothetical constructs at all (i.e. it is not a matter of relevance *tout court*), but instead about the *way* in measures of that variable relate to psychological construct validity. An example is the debate about whether double dissociations in neuropsychology—i.e. instances where process A can experience deficits without deficits in process B and vice versa—actually reveal the existence of distinct cognitive components or processing elements (e.g., Plaut 1995, Patterson and Plaut 2009). While some researchers hold double dissociations as the gold standard for individuating cognitive processes (e.g., Shallice 1988), others (e.g., Plaut and Patterson 2009) argue that apparent double dissociations can in fact arise from lesioning a single cognitive system, and are therefore not strong evidence for the existence of multiple systems or cognitive components. The worry here is not that dissociations are unrelated to cognitive architecture, but that one cannot read architecture directly from dissociability. Finally, there are questions about the statistical techniques used to compare operationalize measures.

Historically, psychologists have often developed new experimental methods for studying mental processes. For example, Vigouroux (1879) introduced the galvanic skin response as a means of measuring emotional distress. These techniques introduce new dependent variables (e.g., reaction times, galvanic skin responses, BOLD responses) into psychology. At the same time, they raise questions about the relevance (Does infant gaze time have anything to do with their “expectations” about the world?) and inferential validity (Do double dissociations reveal the existence of distinct cognitive components?) of psychological experiments employing these

techniques. The question about fMRI and cognitive ontologies, then, is whether experiments that use the BOLD signal as a dependent variable bear on psychological construct validity.

### 3.3 fMRI AND THE CASE FOR REVISION

#### 3.3.1 The call for revision

Many prominent neuroscientists argue that fMRI necessitates cognitive ontology revision. In addition to those who think that neuroimaging studies can be used to discover novel psychological constructs (Price and Friston 2005, Anderson 2010, 2014), there are neuroscientists who believe that fMRI studies can motivate the lumping (e.g., De Brigard et al. 2013), splitting (Lenartowicz et al. 2010, Woo et al. 2014), or elimination (e.g., Greene and Haidt 2002, Lindquist et al. 2012) of psychological kinds. This movement has precursors in philosophical arguments that psychological and/or folk psychological kinds will be radically revised—e.g., reduced (Bickle 1995) or eliminated (Churchland 1981)—as brain research progresses. The core claim is that neuroimaging studies, like behavioral or neuropsychological studies, bears on the validity of psychological constructs. The question is: Can fMRI research play this role in the development and assessment of cognitive theories?

Let me put my cards on the table before forging ahead. Some philosophers and cognitive scientists are wholesale skeptics concerning the *very relevance* of fMRI for cognitive theorizing. Fodor (1999) and Coltheart (2004, 2006) argue that fMRI only reveals *where* cognitive processing takes place without saying anything about *how* it takes place. Furthermore, since cognitive theories do not make predictions about patterns of brain activation, brain activation

patterns (whether measured through fMRI, EEG, or other technologies), cannot contradict cognitive theories, and therefore cannot be used to assess them. I think these arguments are mistaken.<sup>28</sup> As many theorists have recently argued, cognitive theories do make predictions about patterns of brain given a set of auxiliary assumptions (Roskies 2009), neural embellishments (Coltheart 2013) or associative bridge laws (Nathan and Del Pinal 2016) specifying how brain activity relates to cognitive processing. These assumptions, which add content not specified in, say a strictly mathematical or computational description of a psychological task, permit inferences from brain activation patterns to psychological theories. For example, let's suppose that neuroscientists can establish that brain pattern P in the auditory cortex reflects the presence of inner speech (this is a specific brain mapping). In this case, researchers can provide inductive support that a new task—e.g., solving a math problem— involves inner speech based on whether that task elicits pattern P.

As we will see, this means that the value of fMRI for assessing the validity of psychological constructs depends on the plausibility of the relevant bridging assumptions. One consequence of this is that there is no *a priori* reason why fMRI data cannot bear on the validity of psychological constructs. On the other hand, one should not take for granted that neuroimaging data will be useful for assessing construct validity given this reliance on often untested bridging assumptions. For example, if it turns out that the brain exhibits massive degeneracy (Price and Friston 2002)—i.e. if many different regions perform the same function in different contexts—activation in two brain areas wouldn't reflect the recruitment of distinct cognitive processes. My goal is to clarify what it would take for fMRI studies to necessitate cognitive ontology revision and to determine whether fMRI studies *currently* provide strong

---

<sup>28</sup> I am also not convinced that general arguments for the autonomy of psychology as a special science guarantee that brain data could not compel a revision of cognitive kinds, but I will not press the issue here.

evidence in favor of lumping, splitting, or eliminating cognitive kinds. This requires critiquing extant evidence for fMRI-based cognitive ontology revision. A couple methodological notes are in order. First, many of the inferences I identify below are more general than the specific mapping above. My claim is not that, in principle, one cannot provide inductive support for a brain mapping that will lend itself to using fMRI to test some cognitive theory. Instead, my claim is that many prominent studies calling for cognitive ontology revision rely on general bridging assumptions that are at best untested and at worst likely false.

### **3.3.2 Evidence for revision**

Arguments for fMRI-based cognitive ontology revision generally take the following form: (1) There is a mismatch between BOLD data and what our psychological taxonomies, plus some assumptions about how cognitive functions map onto the brain, would predict, (2) It is more likely that our psychological theories are impoverished than that our assumptions about the brain are mistaken, (3) Therefore, we should revise our cognitive ontology to eliminate the mismatch. As I demonstrate below, (2) often involves assumptions about how cognitive functions generally map onto the brain, plus the implicit belief that these bridging assumptions are more secure than one's cognitive ontology. In other words, researchers frequently assume a general picture of how cognitive functions map onto neural structure, and use this general picture to motivate revision. Here is a brief summary of recent studies advocating cognitive ontology revision on the basis of fMRI data.

### *Neural Overlap and Kind Lumping*

Some neuroscientists argue that neural overlap between regions supporting distinct cognitive functions provides evidence in favor of lumping those functions together. De Brigard and colleagues (2013) for example, argue that the same network of brain regions is recruited for episodic memory (thinking about one's past experiences) and other functions such as future thinking and counterfactual thinking about the self. They claim that this network implements a generally mechanism for self-projection and speculate that memory is perhaps not a separate capacity from counterfactual thinking about personal episodes. Lenartowicz and colleagues (2010) argue that a pattern classifier fails to distinguish brain scans of participants engaged in cognitive control processes such as "task-switching" versus "response inhibition;" thus it is likely that these are different verbal labels for the same cognitive control construct. Finally, Eisenberger, Lieberman and Williams (2003) argue that social exclusion (e.g., looking at a photograph of a lover who rejected you) recruits many of the regions involved in somatic pain, thus social rejection involves the experience of physical pain (see also MacDonald and Leary 2005).

### *Neural Dissociations and Kind Splitting*

Other neuroscientists argue for splitting psychological kinds on the basis of dissociations in patterns of brain activity. Contrary to Eisenberger, Lieberman and Williams (2003), Woo et al. (2014) report that, despite some regional overlap between pain and social rejection, a pattern classifier can robustly discriminate neural signals related to pain and social rejection. They conclude that social rejection does not recruit the same brain circuits as somatic pain, and is therefore not literally a form of pain. Epstein (2008) proposes that that visual navigation consists

of both the construction of visual scenes as landmarks and a mechanism that places those scenes in a map-like representation of the environment. Epstein argues for these components based on a neural dissociation between the parahippocampal place area (PPA), which is involved in scene construction, and retrosplenial cortex (RSC), which is involved situating local scenes in an environmental context. Henson et al. (1999) tested a dual-process theory of recognition memory using fMRI. They compared items in a recognition task that participants reported “remembering,” (recollection) versus “just knowing” (familiarity). They found that a portion of anterior cingulate cortex was more active for recollection, while a region in right lateral frontal cortex was more active for familiarity. They conclude that recognition memory consists of at least two different processes.

### *Regional Selectivity and Elimination*

Some neuroscientists draw inferences from regional specialization—i.e. whether or not some brain area is dedicated to a particular task or domain—to cognitive specialization. Lindquist and colleagues (2012) argue that while basic emotion theory predicts that each emotion construct—anger, disgust, etc.—should be associated with a neural substrate specialized for that emotion category, each limbic region is in fact recruited for many different emotion categories. They conclude that the brain does not respect basic emotion theory (it does not exhibit systems dedicated to discrete basic emotion categories), and therefore basic emotions such as “sadness” are not valid psychological constructs. Similarly, Greene and Haidt (2002) argue that the brain areas recruited for moral judgments are also recruited in a variety of other social (e.g., theory of mind) or cognitive (e.g., working memory) processes. In short, there is no special part of the brain devoted to moral reasoning. They take this as evidence that humans do



not possess a distinct moral faculty, but instead moral judgments arise from domain general cognitive, social, and affective processes.

### **3.3.3 Bridging assumptions in fMRI studies**

Each of these studies takes an apparent mismatch between fMRI data and what psychological theory would predict (given some view of how psychological constructs map onto the working parts of the brain) as evidence of a need to revise one's cognitive ontology. For instance, the idea that moral reasoning depends on a special set of moral rules (Mikhail 2007) seems at odds with the fact that the brain regions recruited for moral judgments (assuming the scenarios present ecologically valid instances of moral judgments) are just different permutations of areas involved in garden variety emotional, cognitive, and social functions. This line of thinking suggests there are no special moral rules since there is no special part of the brain that houses them. But what does it take for fMRI data to provide evidence for or against some cognitive theory? Can neuroscientists really look at a brain scan and conclude that some cognitive theory, such as basic emotion theory, or moral grammar theory, is mistaken about the functional components of the mind?

As I mentioned above, some researchers are skeptical about the very notion that fMRI, or any neuroimaging technique, could tell us something about how the mind works (Fodor 1999, Coltheart 2004, 2006). According to this way of thinking, psychological theories do not make predictions about brain activity patterns, so brain activity patterns cannot conflict with them. My reply follows a line of thinking proposed by Roskies (2009, see also Nathan and Del Pinal 2016): Indeed, psychological theories alone—e.g., boxological models or computational models pitched

in mathematical terms—do not make predictions about patterns of brain activations. The ability to link BOLD activation patterns to cognitive theories depends on a set of auxiliary hypotheses or bridging assumptions specifying how neural tissue implements cognitive functions. A brief example will illustrate this point.

Suppose a neuroscientist wants to know if top-down attention and working memory are the same cognitive process. A series of fMRI studies reveal that both working memory and top-down attention robustly elicit activation in DLPFC. For simplicity's sake, assume that DLPFC is the only region differentially activated by top-down attention and working memory tasks. Does this mean that working memory and top-down attention are the same cognitive process? The answer depends on facts about the brain. If each brain area performs a single function, then it is likely that “top-down attention” and “working memory” are different terms for the same cognitive function. If however, brain areas have radically context-sensitive functions—i.e. they perform different mechanistic or computational operations—then it remains possible that “top-down attention” and “working memory” denote quite different cognitive functions that happen to be carried out by the same brain area. In short, the idea that the anatomical modularity of the brain mirrors the functional modularity of our cognitive architecture—i.e. that *each brain region carries out one unique functional component of cognition*—acts as a bridging assumption that allows neuroimaging results to speak to cognitive constructs (see also Bergeron 2007).

Thus Fodor (1999) and Coltheart (2006) are simply wrong to think that fMRI results *cannot* inform cognitive theorizing. However, the value of fMRI for testing cognitive ontologies depends on the bridging assumptions used to link fMRI findings to psychological kinds. And there is no guarantee that these bridging assumptions are true—for example, in [Chapter 2](#), I showed that the assumption that the brain's anatomical modularity corresponds neatly to the

functional architecture of cognition is likely unwarranted. In fact, many of the bridging assumptions that neuroimagers invoke are largely unexplicated or, when they are, untested.

In what follows, I demonstrate that: (1) Cognitive inferences in different fMRI experiments rely on different bridging assumptions, (2) These assumptions often hinge on facts about the brain's functional topography, and (3) These assumptions vary in both their strength (i.e. in extent to which they make commitments about structure function mappings) and their plausibility (i.e. the extent to which the assumption is likely given what we know about the brain). In other words, while we may lack ground truth about the status of many bridging assumptions, we can assess their credibility given what we know now (see Roskies 2009 for an optimistic picture, see Coltheart 2013 for a pessimistic one). Then I argue that, in fact, many of the bridging assumptions currently used to link fMRI data to cognitive ontologies are implausible or overly strong given the state of human brain mapping.

### **3.4 A BRIDGE TOO FAR? FAILURES OF BRIDGING ASSUMPTIONS**

#### **3.4.1 Neural reuse and dedicated substrates**

As I discussed in [Chapter 2](#), a plausible view of the brain's functional topography—in light of the multi-functionality, and hence low selectivity, of individual brain areas—is a picture in which: (1) Many cognitive functions will map onto large-scale brain networks instead of individual regions, and (2) The component operations  $\Phi$ s performed by individual regions or mid-scale brain networks are extensively re-deployed for different cognitive capacities  $\psi$ s. While many philosophers (Lloyd 2000, Bergeron 2007) and cognitive scientists (McIntosh 2000, Uttal

2001, Anderson 2010) advocate this picture as a corrective to what they perceive as an overly modular, region-focused view of brain mapping—offenders frequently include Kanwisher et al. (1997), Saxe and Kanwisher (2003), Kanwisher (2010), etc.—the *consequences* of accepting such a view for psychological inferences in fMRI are rarely discussed in detail (but see Poldrack 2006, Anderson 2010).

Take Greene and Haidt's (2002) assertion that since fMRI reveals no distinctly “moral” centers of the brain, moral judgment is not a unique psychological capacity. This inference relies on a bridging assumption that *each psychological capacity has a unique or dedicated neural substrate*. In light of recent developments in brain mapping, this assumption is misguided. Given the high degree of multi-functionality at the level of individual regions, there will be significant overlap between the functional networks that carry out different cognitive functions. If the same set of brain regions can combine, and recombine, to produce different psychological capacities, then it may be rare to find regions dedicated to a single cognitive function.

Or consider Eisenberger, Lieberman, and Williams' (2003) claim that pain and social exclusion many of the same brain areas—e.g., dorsal anterior cingulate cortex and anterior insula—and thus we should understand social exclusion as a manifestation of somatic pain. This inference relies on a bridging assumption that *neural overlap indicates psychological overlap*. Put differently, shared neural substrates indicate shared psychological operations. This bridging principle is weaker than the assumption of dedicated substrates because it allows for the possibility that psychological capacities can share component operations. Nevertheless, the assumption is still overly strong given the degree of neural reuse observed in human neocortex (Poldrack 2006, Anderson 2010, 2014).

First, it is possible that some brain areas perform different computations, or implement different operations, in different neural contexts. In this case, a region might perform quite different functions when recruited for different contexts. But even if we suppose that each region performs the same operation  $\Phi$  in different contexts ( $C_1$ ,  $C_2$ ,  $C_3$ , etc.), this does not guarantee that neural overlap (e.g., psychological capacities sharing one or more region in common) corresponds to the kind of psychological similarity that researchers often imagine. My concern is that neuroscientists often take neural overlap to indicate similarity at the level of capacities, rather than constituent roles.

Neural reuse forces psychologists to confront the fact that the operations  $\Phi$ s composing a capacity of interest may be quite different than the capacities themselves. Tettamanti et al. (2009) argue that Broca's area is recruited for both visual pattern learning and speech production. This means, *ipso facto*, that pattern learning and speech production share a degree of neural overlap. Nevertheless, visual pattern learning and speech production are, in many senses, different functions. One involves motor control; the other is visual. One involves learning new patterns; the other involves using stored pattern to execute motor commands. Tettamanti et al.'s (2009) claim that Broca's area is for "syntax without language" is simply that learning complex visual patterns involves some of the same computations deployed in producing speech.

Bechtel (2005) argues that a tendency to *elide capacities and operations* during functional analysis occurs in other domains of science. According to Bechtel, biochemists working on the mechanisms of fermentation kept trying to analyze the constituent chemical operations as fermentations themselves.<sup>29</sup> For example, they would ask of *intermediate stage* of the process whether it would ferment as rapidly as sugar. It was not until organic chemistry

---

<sup>29</sup> For a detailed description of the historical case, see Bechtel (2005, 318).

provided a general set operations—e.g., phosphorylation—involving the addition or subtraction of functional groups, that they began to make significant progress in understanding the mechanism of fermentation. In these cases, scientists use functional vocabulary appropriate to the functioning of a whole mechanism (fermentation), and apply it to constituent operations of the system (e.g., the set of reactions involved in fermenting sugar into alcohol).

There is a similar tendency, I think, for cognitive neuroscientists to take neural overlap to indicate functional similarity at the level of capacities rather than the level of roles or operations. For example, neuroscientists will see an area involved in social cognition deployed in a general cognitive task and suppose that the task itself is somehow social. Eisenberger, Lieberman, and Williams (2003) claim that the recruitment of regions like the dorsal anterior cingulate cortex and the anterior insula in both pain and social exclusion suggests that social exclusion is a form of somatic pain. But both of these regions are implicated in a host of functions (Menon and Uddin 2010)—that have nothing to do with pain experience per se. Instead, they may have to do with aversion/avoidance, attentional modulation, or even the detection of salient environmental stimuli. Neural overlap can indicate that two processes have similar downstream consequences—e.g., perhaps pain, social rejection, and other functions all elicit avoidance responses—or indicate shared operations that bear little resemblance to the capacities of interest.

Finally, some neuroscientists argue that neural dissociations—i.e. differences in patterns of BOLD activity—can be used to motivate splitting cognitive kinds. For example, Henson et al, (1999) use BOLD dissociations to motivate a dual-process account of recognition memory in which “recollection” and “familiarity” are understood as distinct processes. This stands in contrast to theorists who see both knowing and remembering as the same basic process operating at different strengths. These inferences rely on the bridging assumption if *two tasks elicit*

*activation in different parts of the brain, those tasks likely recruit different cognitive processes.*

In a sense, this is a fairly weak bridging assumption since it mainly relies on the fact that different brain regions perform different functions. One problem for this assumption would be if the brain exhibits a high degree of *degeneracy*—i.e. if many different brain regions subserve the same basic cognitive function in different circumstances (Figdor 2010, Price and Friston 2002). Another problem is that the same distributed network may carry out both functions, but exhibit different activation peaks in each case. Despite these problems, I think that if two kinds of tasks robustly exhibit activation in non-overlapping brain areas, and the regions in question are not connected to one another; this provides some evidence that the tasks tap into different cognitive processes. The problem is that many tasks of interest exhibit somewhat overlapping patterns of BOLD activity.

Certainly, many of the concerns I have just laid out will be susceptible to a challenge from best practices. For example, anatomical findings can shed light on whether two tasks with different activation peaks are plausible part of the same anatomical network. Likewise, it is possible for neuroscientists to avoid importing the language of capacities when describing shared neural substrates. For instance, Aminoff, Kveraga, and Barr (2013) do not take the recruitment of parahippocampal cortex for not just scene processing, but also spatial navigation, and odor-odor pairings to indicate that odor-odor pairings are processes “visually;” instead they posit a shared component operation. Nevertheless, many of the problematic assumptions I identified are routinely used in drawing cognitive inferences from fMRI experiments. Furthermore, in many cases—e.g., the degree to which the brain exhibits degeneracy, or the degree to which the functions of individual brain regions are context-sensitive—it is unknown whether the bridging assumption is a good one. In the following section, I argue that this general problem is

compounded by the fact that in many cases, it is difficult to even specify the neural predictions a cognitive theory should make (to say nothing about whether that or not that prediction relies on warranted assumptions about neural tissue). This is especially true of cognitive questions that go beyond “counting” the number of processes present or absent in performing a task. I elaborate on this idea using the case of basic emotion theory.

### **3.4.2 A case: fMRI and basic emotion theory**

Having illustrated some of the challenges associated with fMRI-based cognitive ontology revision, I now turn to the case of basic emotion theory to further analyze the challenges associated with bringing fMRI-based cognitive ontology revision. In particular, I focus on Lindquist and colleagues’ (2012) claim that the brain does not respect basic emotion categories, and therefore emotions such as “anger” or “disgust” are illegitimate psychological constructs. My main point here is that using fMRI to assess construct validity depends not only on general assumptions about how cognitive functions map onto the brain—e.g., *most cognitive functions are carried out by individual brain regions*—but also on bridging assumptions that articulate the neural predictions of *specific cognitive theories*. Coltheart (2013) calls these assumptions “neural embellishments.” There are often deep disagreements about the empirical predictions of cognitive theories, and these issues apply to fMRI just as they do to other experimental paradigms (e.g., behavioral measures or neuropsychological dissociations).

Basic emotion theory—and influential theory of the emotions associated with Charles Darwin (1872), Paul Ekman (1973), and others—proposes that emotions fall into discrete categories such as disgust, anger, sadness, happiness, fear, and surprise. A central tenet of basic



emotion theory is that humans, owing to their common evolutionary lineage, contain a universal set of core “affect programs” consisting of stereotyped expressive, physiological, neurobiological, and behavioral components (Ekman 1999, Izard 2011). According to basic emotion theorists, a key prediction of the theory is that basic emotions, such as anger and sadness, have “unique physiological and neural profiles” (Hamann 2012). In other words, basic emotion theory predicts that the brain has a system devoted to anger, a system devoted to disgust, etc. Presumably, the same core systems are active during different bouts of these emotions. By contrast, social constructionists (Lindquist and Barrett 2008) hold that emotions arise from different mixtures of a common set of cognitive (e.g., evaluative), physiological, and social-cognitive elements, such that: (1) The brain has no identifiable system for happiness, fear, etc., and (2) Different instances of what we call “fear,” would consist of different neural and psychological ingredients.

Thus *basic emotion theory predicts that the brain has distinct, consistently identifiable systems for basic emotion categories*—e.g., the brain has an anger system, a disgust system, etc. The amygdala, for example, is usually associated with fear. Not only do fMRI studies of fear typically recruit the amygdala, but patients with amygdala lesions also experience decreased fear responses. One patient with bilateral amygdala lesions was able to withstand startling, proximity to snakes and spiders, and other fearful stimuli, without exhibiting a response (Feinstein et al. 2011). Similarly, the insula is associated with disgust. Patients with damage to the anterior insula often report reduced disgust in response to pictures of common elicitors (e.g., feces, spoiled food, or mutilated bodies) and have difficulties recognizing disgusted reactions on others’ faces (Calder et al. 2000, Kipps et al. 2007). Basic emotion theorists (e.g., Ekman 1999) argue that the

*selective association* of distinct brain regions (e.g., anterior insula) with distinct emotion categories (e.g., disgust) supports basic emotion theory.

Lindquist et al., however, claim that fMRI findings contradict the neuroscientific predictions of basic emotion theory. They performed the largest meta-analysis of fMRI studies of emotion to date. They identified a set of regions consistently activated in studies of emotion (e.g., insula, orbitofrontal cortex, amygdala, etc.) and tested the extent to which each region was associated with a single basic emotion category (e.g., anger or fear). Their main finding was that *every region* in the set is recruited for multiple basic emotion categories. For example, the amygdala is recruited not only for fear, but also anger. The anterior insula is recruited not just for disgust, but also sadness. This lack of specificity held for every cortical and limbic structure in the analysis. Contrary to what basic emotion theory predicts, the brain does not seem to possess systems dedicated to particular basic emotion categories. Lindquist et al. conclude that the brain does not respect basic emotion categories, and thus anger, disgust, etc. are illegitimate psychological constructs that are ripe for elimination.

A shortcoming of Lindquist et al.'s approach is that only looks for specificity at the level of areas (Scarantino 2012, Hamann 2012). That is, their findings are consistent with the possibility that basic emotions are associated with *distributed functional networks* rather than *individual regions*. As we have seen, the brain might contain networks devoted to particular functions—e.g., a network for word form recognition and a network for tactile object recognition—that exhibit some degree of regional overlap—e.g., in posterior lateral fusiform gyrus. In a similar fashion, basic emotion categories may correspond to large-scale brain networks that contain some overlapping regions. As Scarantino (2012) notes, Lindquist et al. seem to embrace a “radical localizationism” in which basic emotion theory predicts not just that

the brain has a disgust system, but that this system is localizable to some particular region. This seems like an overly strong bridging assumption.

Why do Lindquist et al. (2012) contend that a lack of specificity at the level of individual regions is troubling for basic emotion theory? One might think that they simply neglect the possibility of specificity at the level of networks. However, Lindquist et al. (2012) acknowledge this possibility, writing, “it remains a possibility that we failed to locate a brain basis for specific emotion categories because emotion categories are represented as [distinct] anatomical *networks* of brain regions” (2012, 141). As it turns out, recent studies support this basic picture. Saarimäki and colleagues (2015), for example, report MVPA results supporting distinct networks for distinct basic emotion categories. Consistent with Lindquist et al.’s (2012) meta-analysis, these networks overlap, particularly in limbic structures. But while they are open to the possibility that basic emotion categories map onto networks rather than individual regions, it is clear that Lindquist and colleagues consider neural overlap between basic emotion systems (even if emotions map onto large-scale networks) problematic for basic emotion theory. According to this line of thinking, basic emotions cannot be “psychologically basic” if they are built up out of psychological components that are shared between emotion categories (see Scarantino and Griffiths 2011). Their claim is therefore not that basic emotion categories must map onto individual regions, but that they must have distinct neural substrates.

The key question becomes: Is basic emotion theory consistent with some degree of neural and hence functional overlap between emotion categories? While some basic emotion theorists (e.g., Ekman and Cordaro 2011) seem to advocate wholly distinct neural systems (Lindquist et al. 2012 obviously agree), I think this requirement is too strong. Basic emotion theory (Ekman 1973) is certainly committed to distinct neural and physiological processes for different basic

emotion categories—e.g., disgust should affect the viscera in a certain way, be associated with a certain facial expression, etc. But this basic picture is compatible with some degree of functional overlap between basic emotion categories. Consider a region involved in the control of a particular facial muscle. Even if anger and disgust involve different facial expressions, these movements might involve some of the same muscle groups (Ekman and Friesen 1976), and hence both recruit this region. This is equally true of the physiological changes involved in emotional expression. Anger and fear may both elicit activation in the amygdala because they both involve sympathetic nervous system activation (Charney and Deutch 1996).

The issue is thus not whether basic emotion categories will share some degree of neural overlap (they will), but whether this overlap suggests that emotional experiences are cobbled together from a host of evaluative and social-cognitive processes, perhaps in a heterogeneous fashion—i.e. one in which different bouts of “disgust” do not involve the same component processes. By itself, the fact that basic emotion categories seem to map onto large-scale networks with overlapping components does not violate the predictions of basic emotion theory, or support a social constructionist account. Once again, we are in a position where general assumptions about how cognitive functions map onto the brain, or about the neuroscientific predictions of particular cognitive theories, are not enough to bring fMRI to bear on cognitive ontologies. What is needed is a deeper understanding of the precise operations those regions are performing.

### **3.5 PATTERN CLASSIFICATION AND COGNITIVE ONTOLOGIES**

So far, I have shown that many cases of fMRI-based cognitive ontology revision rely on problematic bridging assumptions about how cognitive functions map onto the brain—e.g.,

*neural overlap reflects psychological overlap* or *each cognitive function has a distinct neural substrate*. Neuroscientists increasingly report that cognitive capacities map onto large-scale brain networks (Mesulam 1990, McIntosh 2000), and that individual regions are recycled in many different capacities (Poldrack 2006, Anderson 2010). While most neuroscientists acknowledge these changes, the extent to which they challenge standard cognitive inferences in fMRI is often underappreciated. How should neuroscientists respond to these difficulties? One option is to embark on the project of functional revision described in [Chapter 2](#)—i.e. to reconceptualize the functions performed by individual brain regions. A second option is to embrace new techniques for judging similarities and differences in patterns of brain activation.

Recently, a number of neuroscientists have argued that MVPA and related methods can be used to identify the presence of distinct cognitive processes, and is therefore useful for validating and invalidating psychological constructs (Lenartowicz et al. 2010, Poldrack 2010, Woo et al. 2014, Saarimäki et al. 2015). This fairly new strategy relies on the fairly modest assumption that different cognitive constructs—e.g., anger versus disgust—elicit different *patterns* of brain activation that machine-learning algorithms can detect. According to this view, discrimination warrants cognitive kind splitting. In this section, I critique the use of MVPA and other pattern classification techniques for cognitive ontology revision. I claim that successful discrimination is insufficient for compelling kind splitting. I argue that in many cases, classification success may be driven by differences that are *irrelevant* from the standpoint of the cognitive constructs one is hoping to test. I call these potential confounds “additional constructs” and “implementation differences.”

### 3.5.1 Classifier performance and cognitive constructs

Before delving into the issues surrounding MVPA and cognitive ontology revision, a brief note about pattern classifiers in fMRI research. In standard fMRI research, BOLD associations (response similarities) and dissociations (response differences) are often taken as evidence for the recruitment of similar or distinct cognitive processes (Henson 2006, Machery 2012). This can be done in qualitative or quantitative ways. However, there are limitations to what BOLD differences traditional univariate techniques, which rely on signal changes in individual voxels or ROIs, can detect. For example, some functions are co-localized in the same anatomical regions. That is, some signal differences are expressed not in *what* voxels exhibit differential activity, but in the *way* in which those voxels are differentially active. For example, Tasks A and B might both elicit significant activation in voxels 1, 2, and 3 relative to a matched control, but in Task A, voxel 1 is more active while voxels 2 and 3 are less active, while in Task B, voxel 1 is less active while voxels 2 and 3 are more active.

In these cases, MVPA and other pattern classifier techniques, which can detect *patterned differences* across sets of voxels or ROIs, are more sensitive than traditional univariate analyses (Haxby et al. 2001, Norman et al. 2006). In a standard MVPA design, researchers train a machine-learning algorithm (e.g., a linear Support Vector Machine) on sets of brain scans (e.g., dogs versus cats, or happy brains versus sad ones), and then use the classifier to predict which category a novel, untrained scan belongs to. Classification success is often measured in terms of the percentage (0-100) of novel scans the classifier is able to correctly discriminate. Many authors (Lenartowicz et al. 2010, Poldrack 2010) argue that since different cognitive constructs will elicit different patterns of brain activity, successful discrimination can indicate the presence

of distinct cognitive kinds. The basic intuition is that cognitive similarity will be reflected in the similarity of BOLD activation patterns.

According to this line of thinking, MVPA can be used to justify cognitive kind splitting. Woo et al. (2014) argue that while univariate analyses (e.g., Eisenberger, Lieberman, and Williams 2003) show that pain and social exclusion recruit many of the same regions, MVPA can be used to reliably discriminate pain from social rejection. Thus social exclusion is not a variant of somatic pain. Likewise, Saarimäki et al. 2015 claim that while individual regions are not specific for basic emotion categories, activation patterns across distributed sets of voxels can be used to discriminate anger from fear, fear from disgust, etc. This suggests that MVPA and other pattern classifiers may be more useful for carving the brain at its functional joints than traditional univariate analyses.

Lenartowicz et al. (2010) explicitly set out to validate a taxonomy of cognitive control constructs using a pattern classifier.<sup>30</sup> “Cognitive control” is a heterogeneous category of psychological constructs related to executing goal-directed behaviors and exerting control over competing behavioral responses. Cognitive control consists of many different hypothetical constructs—e.g., working memory, task switching, response inhibition, and response selection—that are operationalized using characteristic tasks (e.g., the Trailmaking Task is widely used to measure task switching). Lenartowicz et al. reasoned that: “these constructs [e.g., task switching] should show a good deal of selectivity in their neural representations if they capture distinct components of cognitive control” (2010, 683). For example, if two tasks both allegedly measure response inhibition, those tasks should elicit similar patterns of BOLD activity. Likewise, if two tasks allegedly measure different constructs, they should be highly discriminable.

---

<sup>30</sup> Lenartowicz et al. (2010) used text mining to obtain fMRI scans in BrainMap (Laird, Lancaster, and Fox 2005) associated with Sabb et al.’s (2008) ontology of cognitive control constructs.

Lenartowicz et al. measured the extent to which distinct cognitive control constructs could be discriminated from one another using a pattern classifier applied to BOLD data.<sup>31</sup> They tested each cognitive control construct pair (e.g., task switching versus response inhibition, response inhibition versus working memory, etc.) and, as a control, tested each cognitive control construct against a language construct. First, they found that each cognitive control construct could be distinguished from the language construct. Then, they found that while some cognitive control constructs (e.g., working memory and response selection) could be distinguished from each other, others (e.g., response inhibition and task switching) could not be reliably distinguished from each other.

For Lenartowicz et al. (2010, see also Poldrack 2010) these findings present a kind of “proof of concept” that fMRI-based pattern classification is useful for validating cognitive ontologies. The fact that each control constructs could be distinguished from a linguistic construct, but some cognitive control constructs were confused with each other, provides *prima facie* evidence that similarities and differences in BOLD activation reflect the relatedness of cognitive kinds.<sup>32</sup> Lenartowicz et al. argue that constructs that can reliably be discriminated from each other—e.g., response selection and working memory—are good candidates for distinct psychological kinds, while ones that cannot be discriminated successfully—e.g., response inhibition and task switching—may be different verbal labels for the same underlying construct.

This study represents an important turn in the literature on fMRI-based cognitive ontology revision. If distinct cognitive kinds elicit distinct patterns of BOLD activation, then

---

<sup>31</sup> For a full description of the classification method, see Lenartowicz et al. 2010.

<sup>32</sup> Of course, demonstrating that cognitive control constructs form a “family” of psychological kinds would require testing discriminability against a wide range of other cognitive kinds (e.g., perceptual, linguistic, memory-related, etc.) rather than a single, randomly chosen construct.



neuroscientists can test their cognitive ontologies against fMRI data.<sup>33</sup> Lenartowicz et al. make two claims. First, they argue that “brain activation in conjunction with the label RS [response selection] was distinguishable from that reported in conjunction with CC [cognitive control], RI [response inhibition], or WM [working memory]. Based on this result, we may conclude that RS represents one distinct control function, and thus a distinct entity in the ontology of cognitive control functions” (2010, 687). In other words, the ability to discriminate activation patterns is evidence in favor of kind splitting.<sup>34</sup> Second, they argue “the similarity between [for example] RI [response inhibition] and WM [working memory] may suggest that these constructs share neural systems and thus may be part of the same control function” (2010, 688). In other words, failure of discrimination is evidence in favor of kind lumping.

The main issue raised by this study (and others such as Saarimäki et al. 2015), is whether the ability to discriminate brain activation patterns related to construct labels reveal whether or not those constructs should be distinct items in one’s ontology. One problem is whether a failure to discriminate supports lumping. Another problem is whether successful discrimination supplies evidence in favor of kind splitting. In what follows, I raise some challenges for the claim that discriminability (using a pattern classifier) provides strong evidence in favor of kind splitting. I will briefly note that the main challenge for lumping kinds on the basis of a failure to discriminate is the problem of negative evidence in MVPA. Different classifiers are able to pick up on qualitatively different patterns—for example, a linear pattern classifier can only pick up on linearly separable differences between patterns. Furthermore, training datasets can be richer or

---

<sup>33</sup> Poldrack et al. (2011) have recently developed a “Cognitive Atlas” that extensively outlines taxonomic relationships between cognitive kinds. One of the goals of this atlas is to codify cognitive ontologies to test against fMRI data. The Cognitive Atlas is online at: (<http://www.cognitiveatlas.org>).

<sup>34</sup> Perhaps Lenartowicz et al. (2010) intend to make the weaker claim that discriminability (on the basis of brain activation patterns) is *necessary* for identifying distinct constructs, but not sufficient. However, passages like the one here suggest an inclination toward the stronger sufficiency claim.

more impoverished. This means that whether a classifier can discriminate patterns will depend on many factors related to classifier performance and experimental design.

### **3.5.2 Two confounds: additional constructs and implementation differences**

Neuroscientists want to use successful pattern classification to reflect the recruitment of distinct cognitive constructs—i.e. discriminability warrants kind splitting. But, as Machery (2012) notes, different psychological tasks will usually elicit somewhat different patterns of BOLD activation; the challenge is determining what counts as a *relevant* difference from the standpoint of cognitive theorizing. Unfortunately, many MVPA studies contain sources of irrelevant differences that may be controlled for in traditional designs. These confounds come in two flavors: *additional constructs* and *implementation differences*.

#### *Additional Constructs*

In traditional designs, neuroimagers use cognitive subtraction to isolate BOLD signal related to a construct of interest. Since MVPA studies do not involve cognitive subtraction—i.e. the classifier operates on BOLD data, not a statistical map—researchers often train a pattern classifier on stimuli that differ in both the cognitive construct of interest—e.g., in what emotion is elicited—and in what other processes—e.g., perceptual processes—are present during the scan.<sup>35</sup> Therefore, successful discrimination may be driven by the presence of irrelevant additional cognitive processes. Consider, for example, Saarimäki et al. 2015’s use of MVPA to

---

<sup>35</sup> While it is possible, and often desirable, to select regions of interest for MVPA based on previous subtractions (e.g., one might use a subtraction contrast to generate ROIs for MVPA): (1) Researchers sometimes neglect this issue entirely, opting for whole brain analyses, and (2) Spreading activation may remain a problem since MVPA does not use cognitive subtraction.

discriminate basic emotion categories. Saarimäki et al. used emotional movies—e.g., movies that elicit anger, disgust, etc.—to derive their classifications. But emotional movies do not only differ in terms of the emotions they elicit; they also differ in numerous other factors including visual stimuli (e.g., pleasant ones versus unpleasant ones), auditory stimuli (e.g., high pitched versus low pitched music), etc. Thus there is no guarantee that distinct affective processes drive the classifier's selectivity.

The problem of additional constructs is most salient in studies that use different “delivery methods” for different stimulus categories—e.g., studies which present stimuli via different sensory modalities, or use different tasks as proxies for the same construct. For example, Woo et al. (2014) are interested in comparing neural representations of pain and social exclusion using MVPA. They found that a classifier could distinguish a painful stimulus (caused by intense heat) from a social rejection stimulus (caused by a picture of an ex-partner) 100 percent of the time. Furthermore, a classifier trained to discriminate a neutral warm stimulus from a painful heat was not useful for discriminating a neutral face stimulus from the face of an ex-partner and vice versa. However, these stimuli differ not only in whether or not they both recruit “pain,” but also the intensity of the thermal stimulus, the social context of the encounter, and many other factors. Thus it is not clear that discrimination success is driven solely by whether or not a pattern for somatic pain is present in the stimulus.

It is likely that the presence of additional constructs also confound Lenartowicz et al.'s (2010) use of pattern classifiers to test an ontology of cognitive control. This is because Lenartowicz et al. used heterogeneous sets of tasks as proxies for individual constructs. Neuroscientists use different tasks (the Go/NoGo Task or the Stop-Signal Task) to operationalize the same cognitive construct (response inhibition). But, there is no guarantee that these tasks

recruit one and the same neural or cognitive processes. This problem, widely discussed by Sullivan (2009, 2010) in the philosophical literature, occurs when scientists neglect to test their operational schemes.

Lenartowicz et al. found that response inhibition—a cognitive construct defined as the conjunction of BOLD signal from the Go/NoGo Task, Stop-Signal Task, and Antisaccade Task—could be distinguished from working memory and response selection, but not task switching. One possibility is that response inhibition and task switching are the same construct; another is that both response inhibition and task switching are mixtures of constructs recruited by different tasks. For example, Swick, Ashley, and Turken (2011) performed a meta-analysis of fMRI data for two response inhibition tasks—the Go/NoGo Task and the Stop-Signal Task. They found some degree of BOLD overlap, but also considerable difference, suggesting that while the two tasks *may* share functional components, they should not necessarily be considered “interchangeable measures of response inhibition” (2011, 1662). According to this concern, response inhibition might consist of constructs 1, 2, and 3 recruited by Tasks A, B, and C while task switching consists of 2, 3, and 4 recruited by Tasks D, E, and F. In this case, the success or failure of a classifier would say nothing of interest about the “constructs” one is intending to test. As Figdor (2011) recognizes, we cannot test our ontologies of cognitive kinds without simultaneously testing our ontologies of tasks.

### *Implementation Differences*

Another source of irrelevant differences driving successful fMRI-based classification is what I call “implementation differences.” The general worry is that we have good reason to think that, in some cases, MVPA is sensitive to differences in the *way* a single process is implemented

rather than in *what kind* of process is implemented. In other words, the same system cognitive system can operate on different inputs, which will exhibit somewhat different patterns of neural activity; thus it is difficult to tell which discriminations are due to differences in the kind of process taking place.

In a famous early MVPA study, Haxby et al. (2001) trained a classifier to judge whether a brain scan resulted from a person observing a shoe or a bottle. The classifier was successful at discriminating shoe scans from bottle scans. On the face of it, seeing a shoe is not a different “kind” of cognitive process from seeing a bottle. Both are typically thought of as instances of feeding a general system for visual object recognition (ventral visual areas like lateral occipital complex, inferotemporal cortex, etc.) different inputs. Similarly, one could imagine that a classifier could be trained to discriminate particular memories—e.g., remembering one’s wedding versus one’s dissertation defense—without warranting the claim that different cognitive processes or constructs are involved in constructing these memories.

Philosophically, there are several ways one might want to put this concern. One is that MVPA discriminations might be driven by differences in either vehicles or contents—i.e. in the way that something is represented or in *what* content is represented (see Crane 1995)—but only difference in vehicles warrant positing differences in cognitive kinds. Another that MVPA discriminations might be driven by the same mechanism operating on different inputs—e.g., a camera takes a picture of a bottle versus a shoe—rather than the presence of different kinds of mechanisms, while only the latter would warrant positing different cognitive kinds (Craver 2009).

In both cases, the fact that MVPA can discriminate seeing a shoe from seeing a bottle provides a *reductio ad absurdum* of the idea that it is useful for individuating cognitive kinds.

Lenartowicz et al. (2010) hope that successful classification reflects the recruitment of distinct cognitive processes. But if a sufficiently sensitive classifier can discriminate virtually any two tasks, this undermines the usefulness of discriminations for tracking interesting psychological differences (e.g., between kinds of processes or distinct constructs).

### **3.5.3 Controlling for confounds**

Since discrimination may be driven by implementation differences, or by the presence of additional constructs, one cannot equate “cognitive constructs” with “stimulus classes that can be successfully classified.” This does not mean that MVPA is useless for validating cognitive ontologies. However, it does mean that neuroscientists will need to develop novel experimental controls and ways of integrating pattern classification with other findings. Here I briefly review some options.

One might control for the presence of additional constructs by adopting clever cross-classification designs. Saarimäki et al. (2015) validated their classification of basic emotion categories by testing whether a classifier trained to discriminate emotions on the basis of movies (e.g., sad film clips versus happy ones) could successfully classify scans resulting from internally-generated emotion states (e.g., someone thinking about a sad memory versus a fond one). Likewise, it is possible to evoke the same cognitive process using different stimulus modalities. For example, Damarla, Cherkassky, and Just (2016) found that a pattern classifier trained on representations of quantity (e.g., 3 visually presented dots versus five visually presented dots) was successful at discriminating representations of quantity in other modalities (e.g., hearing 3 tones versus 5 tones). These controls make it more plausible that successful

discrimination is driven by neural representations of quantity rather than confounding factors (see Kaplan, Man, and Greening 2015 for an interesting discussion).

It is harder to control for implementation differences. One possibility is to adopt a compare activation patterns resulting from different processes acting on the same stimuli and/or contents. For example, one could compare activation patterns seeing an apple, seeing a lemon, remembering an apple, and remembering a lemon. The problem here is that discrimination success will not be sufficient to compare these patterns, since it may be possible to achieve high discrimination between all four categories. What is needed is some way of comparing the *similarity* of the patterns used to classify each category; for this, researchers will need to adopt methods such as representational similarity analysis, (Norman et al. 2006). In any cases it is clear that classification alone is not sufficient to warrant cognitive kind splitting.

### 3.6 RETHINKING THE DIRECTION OF REVISION

In this chapter, I argue that fMRI typically only provides weak evidence in favor of cognitive ontology revision. In other words, fMRI is not currently very useful for explicitly testing our cognitive ontology. While the problems I raise are not necessarily in principle problems (and some of them may be corrected by current best practices), they present a cautionary tale for many popular inferences on neuroimaging data. First, I situate the move toward cognitive ontology revision in the context of psychological construct validation. Then I argue that using fMRI to validate or invalidate psychological constructs depends on bridging assumptions about how cognitive functions are instantiated in neural systems. I show that many of the bridging assumptions—e.g., *each cognitive function has a unique neural substrate, neural overlap*

*suggests psychological overlap, each regions performs a single function*, etc.—on which these inferences depend are implausible given current views of the brain’s functional topography (Bergeron 2007, Anderson 2010). Finally, I demonstrate why MVPA is not a panacea for the challenges of inferring cognitive components on the basis of neural associations and dissociations.

The goal of this chapter has been to carve out a philosophical middle ground between the pessimistic conclusion that fMRI is irrelevant for testing our cognitive ontologies (e.g., Coltheart 2006) and the exuberance with which many scientists proclaim that fMRI studies provide one of the strongest forms of evidence available for testing cognitive ontologies (e.g., Lindquist et al. 2012). I have attempted to improve on existing philosophical discussions (e.g., Roskies 2010, Nathan and Del Pinal 2016) by articulating and critiquing many of the bridging assumptions at work in fMRI-based cognitive ontology revision. This discussion raises serious worries about whether cognitive neuroscience is delivering on its promise to revolutionize psychology.

Many philosophers (Churchland 1986, Bickle 1995) and neuroscientists (Poldrack 2010, Anderson 2010) are enamored of the idea that psychology will be radically transformed by its contact with neuroscience. For what it is worth, I too think that insights from brain science are already transforming, and will continue to transform, our psychological taxonomy. Nevertheless, I am concerned that both past and current contributions of neuroscience to the taxonomy of psychology are probably exaggerated.

A familiar story holds that psychologists once thought of “memory” as a unitary psychological kind. Then, neuropsychological studies of H.M. and other amnesiacs with damage to the hippocampus and medial temporal lobes generated a need to split memory into declarative and implicit variants (Scoville and Milner 1957, Milner, Corkin, and Teubner 1968, Cohen and



Squire 1980). Churchland hails this fractionation of memory as case where brain data provoked a revision of psychological categories, writing: “it is this *unprecedented* dissociation of capacities that has moved some neuropsychologists to postulate two memory systems, each with its own physiological basis” (1986, 371, emphasis added). The problem with this story, as Hatfield (2000, see also Frank and Badre 2015) notes, is that it is historically inaccurate. In their major paper splitting memory into declarative and procedural memory systems, Cohen and Squire (1980) traced the distinction between procedural and declarative memory to previous behavioral work, and even to Gilbert Ryle’s (1949) philosophical distinction between “knowing how” and “knowing that.” Indeed, according to Frank and Badre, Corkin (1986) “motivated the investigation of [of HM’s intact motor skills] with ‘observations in normal man’ that motor and other forms of memory were distinct” (2015, 15).

If this historical interpretation is correct, then philosophers have routinely mischaracterized the significance of H.M. for theorizing about the relationship between psychology and neuroscience. This is not a case where brain data led to a revision of psychological categories, but a case in which an existing distinction in cognitive psychology led to a novel characterization of the function of some brain area (e.g., the studies solidified the role of the hippocampus and medial temporal lobes in declarative, but not procedural memory). Indeed, there are numerous cases in which a previously-existing psychological taxonomy—e.g., Baddeley and Hitch’s (1974) working memory model—was used to identify functional brain structures. However, it is harder to find cases in which neuroimaging data compelled a radical revision of psychology.

## 4.0 NETWORK INFERENCES AND THE USES OF RESTING STATE fMRI

### 4.1 CHAPTER 4 INTRODUCTION

If my analysis in [Chapter 3](#) is correct, then neuroscientists frequently use psychological distinctions—e.g., between the central executive and phonological loop components of working memory—to identify functional structures in the macroscale (see Sporns 2011) brain using fMRI. The fusiform face area, for example, was identified as a region responsive to the BOLD contrast between faces and other complex objects. By contrast, it is relatively rare to find cases in which fMRI caused a genuine revision of cognitive theory. While some philosophers (Churchland 1986, Bickle 1995) yearn for brain data to overturn psychology, for the most part, this is not how psychology and neuroscience have historically interacted with each other. Recently, several prominent neuroscientists—e.g., Biswal et al. 2010—have argued that resting state fMRI, a novel way of analyzing fMRI data, can be used to identify functional brain networks without the need for psychological tasks. If this is right, then perhaps one way around the problems I have discussed so far—i.e. problems with specifying the functions of particular brain regions—is to adopt a method that permits researchers to carve the brain at its functional joints without a prior understanding of psychological theory. In this chapter, I examine the use of resting state fMRI to identify large-scale functional brain networks. The main question is: Can fMRI be used to identify functional brain networks in a “bottom up” fashion that does not rely on

prior cognitive theorizing? Note: In this dissertation chapter, I use the term “I,” but this chapter is fully co-authored with David Danks, professor of philosophy at Carnegie Mellon University. We contributed equally to the work, which I have modified to connect with the broader themes of the dissertation.

Traditional fMRI experiments involve manipulating specific stimuli or tasks, and measuring resultant changes in the blood oxygen-dependent (BOLD) signal. In standard subtraction designs, researchers compare activation patterns from a task of interest (e.g., reading words) to patterns in a matched control task (e.g., reading meaningless non-words such as “BLORT”) to isolate task-related BOLD signal changes. In this framework, task manipulations are the independent variables, BOLD changes are the dependent variables, and other sources of BOLD signal fluctuation—e.g., fluctuations due to cardiac cycle, head motion, or unexpected task differences—are treated as sources of noise to be eliminated by sound experimental design or principled statistical adjustments (Purdon and Weisskoff 1998, Aguirre 2014).

Although task-based experiments were the initial focus of fMRI research, neuroimagers soon began to observe a puzzling pattern. Many brain regions seem to exhibit stable, correlated, low frequency ( $\sim 0.01$ - $0.10$  Hz) BOLD fluctuations when participants are “resting” in the scanner magnet between task blocks (Biswal et al. 1995, Binder et al. 1999, Gusnard and Raichle 2001, Raichle et al. 2001). This surprising, accidental finding led to the characterization of the so-called “default mode network,” a set of regions whose activity is correlated at rest, but relatively uncorrelated during specific tasks (Raichle et al. 2001, Greicius et al. 2003). The immediate conclusion was that perhaps some endogenous BOLD fluctuations are not “noise,” at all, but instead reflect previously uncharacterized patterns of brain activity. This insight suggested a paradigm shift towards studying the brain’s intrinsic functional dynamics, where “intrinsic” is to

be understood as spontaneously or internally generated, as opposed to driven by a specific task or stimulus (Snyder and Raichle 2012).

Research on the default mode network has led to the general methodological development of resting state fMRI. Resting state fMRI experiments measure “functional connectivity” patterns in resting participants. That is, they measure BOLD correlations between target voxels or “seed regions” (predefined brain areas) that are thought to reflect network level brain patterns of activity (Power et al. 2014). Resting state fMRI is a burgeoning area of clinical and basic brain research. As Power, Schlaggar, and Petersen (2014, 692) note, “resting state fMRI has grown from an unexpected observation in fMRI ‘noise’ to a major area of human neuroimaging.” While there is no unified goal of resting state research ([more on this in 4.2](#)), the basic idea is that resting state BOLD fluctuations can be used to identify and investigate “functional” brain networks, where here “functional” means, roughly, “performs some metabolic or cognitive function.” Note from the outset that “functional connectivity” is just a fancy way of saying “BOLD correlations;” as we will see, the actual “functional”—i.e. cognitive or metabolic—significance of resting state connectivity networks is often very unclear.<sup>36</sup>

Resting state fMRI is a philosophically important enterprise. First, it represents a departure from the stimulus-response paradigm that characterizes much of cognitive psychology and neuroscience, and thus potentially provides a window on “naturalistic” (i.e., outside of the lab) cognition, or on psychological processes (e.g., mind wandering) that are difficult or impossible to study using explicit task instructions (Snyder and Raichle 2012). Second, resting state fMRI sidesteps one of the main critiques of task-based fMRI. Standard studies almost

---

<sup>36</sup> Throughout this chapter, I will use the terms “functional connectivity” or “connectivity network” to refer to correlations in the BOLD signal and the terms “functional brain network” or “functional network” to refer to networks of the brain that perform some cognitive or metabolic function.

always employ a logic of subtraction, which makes the assumption of “pure insertion” (that researchers can construct task-pairs that differ solely in the recruitment of one cognitive component), an assumption that faces significant problems (Friston et al. 1996, Van Orden and Paap 1997, Uttal 2001, 2008). Resting state fMRI aims to identify functional brain networks without making prior assumptions about the psychological differences between tasks (as there are no experimenter-provided tasks at all), and thus could represent a major methodological advance in human brain mapping. Third, resting state fMRI raises important questions about the value of exploratory experiments (Steinle 1997, Franklin 2005) in cognitive neuroscience. Proponents of resting state studies emphasize that the methodology holds the promise of simultaneously identifying multiple functional brain networks in a “bottom up” fashion, prior to any theorizing about precisely what those networks might do (Biswal et al. 2010, Snyder and Raichle 2012). Thus resting state fMRI may be a useful tool for *discovery* instead of, or in addition to, hypothesis testing (Biswal et al. 2010).

At the same time, and despite the proliferation of resting state fMRI research, there is little consensus about what physiological or cognitive processes resting state analyses are actually measuring. Some researchers (e.g., Andrews-Hanna, Smallwood, and Spreng 2014) give specific cognitive interpretations to resting state connectivity patterns—e.g., the default mode network is claimed to be associated with imagination, self-reflection, or planning—while others claim only that resting state analyses measure some form of endogenous or intrinsic functional activity—e.g., background or “housekeeping” functions that keep the brain ready to adaptively respond to external inputs or internal demands (Raichle 2009, Snyder and Raichle 2012, Klein 2014). Some researchers focus their efforts on regions whose activity is only or primarily correlated at rest (Biswal et al. 1995), while others emphasize similarities in functional

connectivity between tasks and rest (Bray et al. 2015, Cole et al. 2014). Given this plethora of theoretical interpretations, what exactly do resting state analyses tell us about the brain? This question has not received enough critical attention in the philosophical and scientific literature (though see Morcom and Fletcher 2007, Klein 2014).

In this chapter, I examine what resting state analyses do, and do not, reveal about the brain's functional organization. Proponents of the approach frequently assign functional (i.e. cognitive or metabolic) significance to resting state connectivity patterns. I claim that these inferences from statistical structure—i.e. functional connectivity patterns—to functional organization are likely unwarranted, and in doing so provide a novel interpretation of extant findings in resting state research. I argue that resting state analyses involve sampling from a “mixture distribution”—the probability distribution that results from sampling from multiple, distinct, heterogeneous populations without knowing the population from which each individual is drawn (Redner and Walker, 1984, Zheng and Frey 2004). That is, the functional connectivity patterns observed in resting state analyses may arise from temporally and causally distinct processes (e.g., different cognitive processes) that are blended together by current ways of sampling and analyzing resting state datasets. Networks inferred from data from a mixture distribution can be quite different from any particular component network, or even from the superposition of the individual networks. Thus, there is no straightforward way to attribute functions, cognitive or otherwise, to resting state brain networks, precisely because (on this view) there is no single, stable resting state network. In short: the large-scale functional connectivity “networks” (e.g., the default mode network) observed in resting state studies may be sampling artifacts rather than genuine features of the brain's functional organization.

I argue that extant evidence is compatible both with this hypothesis—I refer to it as the “mixture” or “superposition” view—and with the thesis that resting state analyses measure the endogenous activity of a handful of large-scale, functional brain networks. The components of these mixtures may reflect the brain engaging in a series of “hidden tasks” such as imagery, inner speech, or planning, whether instead of or in addition to intrinsic or endogenous activity. Inferences from connectivity patterns to cognitive or neural function are substantially more complicated in resting state fMRI research than is standardly recognized.

The superposition or mixture view has not received much discussion in the existing literature. I aim to remedy this situation and to make three contributions to the theoretical literature on resting state fMRI. First, I clarify different uses of the relatively new technique of resting state fMRI, as well as delineate some open controversies. Second, I aim to demonstrate that the superposition view is both plausible and raises an interpretive problem for the use of resting state analyses in brain mapping. Finally, I argue that accounting for mixtures may involve not only the use of new technical tools, including demixing methods such as changepoint detection algorithms (e.g., Adams and MacKay, 2007; Desobry, Davy, and Doncarli, 2005), but also relaxing the experimental standard that resting state analyses be completely cognitively unconstrained. There may be an interesting tradeoff between purely task-driven and task-free experimental designs. Some of these recommendations are already being implemented independently (e.g., Vemuri and Surampudi 2015), and combining them may significantly improve our ability to identify candidate functional networks using resting state analyses.

[Section 4.2](#) reviews the experimental design of resting state studies, and explains some interpretations of resting state findings. [Section 4.3](#) then presents the mixture or superposition view, and explains the ways in which inferences from structure to function in resting state fMRI

are complicated by the fact that resting state analyses involve sampling from a mixture distribution. In [Section 4.4](#), I address a number of empirical objections to the superposition view, such as the idea that the high degree of intrasubject and intersubject consistency in resting state analyses rules out our interpretation. I argue that our account can withstand each of these objections, and conclude by discussing some methodological and theoretical implications for human brain mapping (Sections [4.5](#) and [4.6](#)).

## 4.2 THE USES OF RESTING STATE fMRI

There is something counterintuitive about the very idea of resting state fMRI. In a traditional framework, the goal of fMRI experiments is to measure task-related changes in the BOLD signal. Under the right conditions, differential BOLD response given experimenter-manipulated tasks is taken to reveal something about where cognitive functions are localized and how these regions respond to different task conditions (Saxe, Brett, and Kanwisher 2006). The goal of these experiments is to tell us what some part of the brain is doing by linking BOLD changes to task manipulations. Even if the assumptions grounding these experiments turn out to be wrong (e.g., Van Orden and Paap 1997), the goal of such experiments is clear.

In contrast, resting state fMRI experiments collect BOLD time series while participants are “resting,” a state operationally defined as lying passively in the scanner without explicit instructions (Snyder and Raichle 2012). Therefore, “the resting state is *uncontrolled* according to the usual conventions that apply to cognitive neuroimaging” (Snyder and Raichle, 2012, emphasis added). What epistemic benefits coincide with the loss of cognitive constraint—i.e. control over what participants are doing—in resting state analyses? In other words, why do



neuroscientists want to study the brain under conditions where they do not know what participants are thinking about?

#### **4.2.1 Resting state analyses**

It is useful to think about resting state fMRI in the broader context of exploratory experiments (Steinle 1997, Franklin 2005, Biswal et al. 2010) that search for meaningful patterns in data without an explicit theory of what gives rise to those patterns. In cellular biology, for example, researchers sometimes measure the mRNA transcripts in a group of cells at various times to determine which gene expression levels correlate with one another. Such an experiment does not involve manipulating an environmental factor and measuring changes in gene expression, but may reveal interesting features of the target system, such as what genes are part of the same regulatory networks, or even aspects of the regulatory network structure (Basso et al. 2005). Similarly, in most resting state fMRI studies, researchers are seeking to learn about activation patterns and networks without necessarily controlling particular input or cognitive tasks. More precisely, given a “resting” participant, neuroscientists define a set of seed regions (or voxels), obtain a BOLD time series for each region, and measure the extent to which those time series are correlated (Power, Schlaggar and Petersen 2014).<sup>37</sup> These correlations in the BOLD signal are known as “functional connectivity” patterns. It is important to note that functional connectivity patterns are strictly correlational, unlike “effective connectivity” patterns, which aim to encode

---

<sup>37</sup> While correlating BOLD time series for predefined seed regions is still the most common way of analyzing resting state fMRI data (Fox, *et al.*, 2005), independent component analysis (ICA) (see Beckman et al. 2005) is an increasingly popular analysis technique. Though many of our concerns apply to ICA as well, I will focus on the traditional way of analyzing resting state data.

causal relationships (Friston 2011). There are limits to what researchers can infer from functional connectivity patterns, though they can rule out many hypotheses.<sup>38</sup>

Data from resting state fMRI experiments are often presented as matrices of Pearson correlation coefficients between seed regions (e.g., a table reflecting how correlated region A is to Region B, Region B is to Region C, etc.) or as graphical networks in which nodes depict regions and edges depict functional connectivity strengths (see Bullmore and Sporns 2009). Strictly speaking, functional connectivity patterns only demonstrate the extent to which changes in blood oxygenation are correlated across the brain. To be interesting from the standpoint of cognition or neurobiology, these BOLD changes should reflect correlated patterns of *neural* activity rather than simply head motion, cardiac cycle, or some other confounding factor (Glover and Lee 1995, Friston et al. 1996, Lowe, Mock, and Sorenson 1998). There is ample evidence that functional connectivity patterns do reflect real neural activity: for example, BOLD fluctuations at rest have been shown to correspond to electrocorticography measurements in surgical patients (e.g., He et al. 2008). Also, several studies indicate that structural and functional connectivity correlate with one another, which is expected if resting state fluctuations reflect neural signaling rather than motion-related or physiological artifacts. For instance, Johnston and colleagues (2008) found that functional connectivity relationships between the hemispheres were intact prior to, but dropped precipitously after, a patient received a complete section of the corpus callosum.

Thus functional connectivity patterns inferred from resting state data likely reveal *something* about the brain's functional—i.e. metabolic and/or cognitive—organization and

---

<sup>38</sup> For example, if Region A and Region B exhibit a high degree of functional connectivity, it could be that activation in A causes activation in B, that activation in B causes activation in A, that both are activated by region C, and so forth.

behavior. The question is: How should neuroimagers *interpret* these patterns? Before considering some of the main uses and interpretations of resting state fMRI, I need to clarify the so-called “resting state” itself. Some authors (e.g., Snyder and Raichle 2012) understand “rest” as a particular kind of brain activity that is intrinsic or internally-driven as opposed to response-driven. According to this view, when participants are not asked to perform a particular task, intrinsic physiological or cognitive processes begin to predominate the BOLD signal at particular timescales. But Morcom and Fletcher (2007) and Klein (2014) both argue that we should be skeptical about inferences from (a) the operational notion of “rest” as undirected time in the scanner, to (b) “rest” as a particular kind of intrinsic brain state. As they argue, there is no guarantee that the brain defaults to intrinsic or internally guided processes simply because participants have no explicit task instructions from the experimenter. Instead, participants may be engaged in a number of “hidden” tasks such as wondering when the session will be over, pondering the weird noises the scanner is making, or planning a trip to Europe. For all we know, “resting” participants may engage in some unknown combination of the cognitive processes (e.g., top-down control of attention, episodic memory, etc.) that experimenters target individually in task-based studies. These critiques suggest that whatever the eventual value of resting state analyses, we ought not *assume* (though it may turn out to be the case) that the “resting state” is a unique kind of brain state.

#### 4.2.2 Interpretations of resting state data

##### *(i) Task-Anticorrelated Networks*

Resting state fMRI is most famously associated with the discovery of the so-called “default mode network,” a network of regions including parts of prefrontal cortex, cingulate cortex, medial temporal lobes, etc. that are more active, and whose activity is more correlated with one another, at rest compared to during a task (Biswal et al. 1995, Greicius et al. 2003; Buckner, Andrews-Hanna, and Schacter 2008). The identification of the default mode network led researchers to theorize that it represents the intrinsic, endogenous, ongoing processes that are not involved in any particular task-based cognition (Shulman et al. 1997, Raichle et al. 2001). For this approach, the main purpose of resting state fMRI is to measure spontaneous activity in regions that are uniquely correlated or active at rest. Analyses of these resting state data are, on this interpretation, thought to reveal endogenous or internally guided neural processes that are the brain’s “default” state of self-reflection, internal monitoring, future thinking, and so forth. However, there is little consensus about what, if anything, is actually being done by this default mode network, even though it has been identified in many studies in fMRI and other imaging modalities such as positron emission tomography (PET) (Gusnard and Raichle 2001, Raichle et al. 2001). Some hypotheses include generating stimulus-independent thoughts, monitoring the external environment, imagination, self-reflection, past and future oriented thinking, and more (Buckner, Andrews-Hanna, and Schacter 2008, Andrews-Hanna, Smallwood, and Spreng 2014).

### *(ii) An Absolute Baseline*

Another reason to perform resting state analyses is to provide an absolute baseline for task-related fMRI (see Gusnard and Raichle 2001). This interpretation holds that physiological baselines obtained using positron emission tomography (PET), can provide an absolute physiological baselines for interpreting task-based changes in PET and fMRI. In standard subtraction designs (see Section 1), neuroimagers compare activity in a task of interest to a matched control. Most people think that relevant brain regions become more active during the task of interest (e.g., reading a meaningful word) compared to a control task (e.g., reading a meaningless non-word), but sometimes, a region is *less* active during the focal task; these are known as “deactivations” in the neuroimaging literature (Gusnard and Raichle 2001). In these cases, it is unknown whether the deactivated area is being actively suppressed (e.g., due to inhibitory influences from other regions), or is merely less active in that condition. According to Gusnard and Raichle (2001), comparing task-related BOLD changes to a baseline obtained at rest can help adjudicate these issues. For example, some regions exhibit deactivation, but are fairly close to their baseline resting levels, while in other cases, a region exhibits deactivation, and is significantly below its resting value. This suggests that perhaps resting state data should be collected and used as a physiological baseline against which to interpret task-related BOLD changes in other experiments.

### *(iii) Correspondences Between Rest and Tasks*

Many studies report similarities between resting and task-evoked functional connectivity patterns (Greicius et al. 2003, Fox and Raichle 2007, Smith et al. 2009, Bray et al. 2015, Cole et al. 2014). For example, the default mode network is implicated in certain tasks including

autobiographical recall (Andrews-Hanna, Smallwood, and Spreng 2014). Smith and colleagues (2009) compared a meta-analysis of task-evoked co-activations (i.e. what regions tend to be activated together in particular tasks) to resting state functional connectivity patterns. They found that regions whose activity is correlated at rest tend to be co-activated in tasks: for example, previously defined visual or motor areas often exhibit a high degree of functional connectivity. Several studies argue that functional connectivity analyses imply the existence of a small number of large-scale intrinsic connectivity or resting state networks (abbreviated ICNs and RSNs, respectively) that correspond to specific cognitive domains—these networks include a dorsal attentional network, a frontoparietal control network, visual networks, motor networks, and more (Damoiseaux et al. 2006, Power et al. 2011, Bray et al. 2015). Crucially, though, these networks are thought to be few in number, and relatively stable over time. For example, they might reflect background activity in the brain’s major functional “pathways.” In general, they are thought to be qualitatively different from the types of networks that are found using standard task-based fMRI experiments.

An exciting example of this approach is Vahdat and colleagues’ (2011) use of resting state fMRI to characterize the functional networks involved in motor learning. They report that, as one might predict, motor learning corresponded to changes in functional connectivity patterns between cerebellar cortex, dorsal premotor cortex, and primary motor cortex. Furthermore, they report that changes in a *novel functional connectivity network* (involving somatosensory cortex, ventral premotor cortex, and supplementary motor cortex) corresponded to perceptual changes accompanying motor learning. These results suggest that resting state fMRI can be used to identify novel functional brain networks—for example, researchers might characterize a new RSN (a set of regions not previously associated with one another) and then search the existing

task-based literature to see what functions the network might perform (Power et al. 2011). These studies reflect a departure from a special interest in “resting” functions to a conception of the field where neuroimagers identify task-related functional networks by probing connectivity in the resting brain. According to this view of resting state fMRI, resting state analyses might reveal candidate functional networks, even if it is unclear what those networks are doing. This would constitute a revolutionary, data-driven “bottom-up” approach to brain mapping, but also assumes that the networks found in these data analyses are relatively stable, which is exactly the target of our critique ([Section 4.3](#)).

*(iv) The Diagnostic Value of Functional Connectivity Structures*

Finally, some uses of resting state fMRI do not involve attribution of specific functions to resting state networks or even the assumption that functional connectivity patterns reflect candidate functional networks. Instead, some researchers use resting state fMRI for purely diagnostic purposes. For example, a number of resting state fMRI studies demonstrate that patients with Tourette’s syndrome (Church et al. 2009), schizophrenia (Arbabshirani, Castro, and Calhoun 2014), or depression (Sheline et al. 2010) exhibit characteristically altered functional connectivity patterns. Similarly, functional connectivity patterns can be used to predict the severity of brain lesions at different sites. For example, Power et al. (2011) report that regions characterized as “hubs” in a functional connectivity network defined using resting state data could reliably predict the severity of neuropsychological deficits resulting from focal lesions to the area. These studies suggest that resting state analyses may have substantial predictive or diagnostic value, even if they have less value for brain mapping or network discovery.

## 4.3 THE CHALLENGE OF MIXTURE DISTRIBUTIONS

### 4.3.1 Functional networks or sampling artifacts?

Despite the different uses and interpretations of resting state fMRI data, there is widespread agreement that it is a valuable technique for identifying large-scale, intrinsically active, candidate functional networks. For instance, Bray and colleagues (2015) argue that resting state analyses routinely converge on large-scale brain networks for attention, vision, self-reflection, motor control, and many more. The inference goes as follows: since the same topographic “structures”—i.e. functional connectivity patterns—occur both at rest and during tasks, this suggests that resting state networks are (1) Coherent functional units that (2) Can sometimes be assigned particular cognitive functions (e.g., attention, self-reflection, or empathy). More colloquially, resting state data can reveal candidate functional networks whose functions can subsequently be attributed by comparing resting state and task-based data (e.g., Vemuri and Surampudi 2015). Even when resting state fMRI analyses identify *novel* candidate functional networks, they are still thought to be part of a small number of large-scale, intrinsically active functional brain networks (Fox et al. 2005, Bray et al. 2015). The field is moving away from the assumption that there are special “resting” functions (e.g., imagination), but still assumes that resting state data reveal a small number of “core” (in some sense) functional networks (Smith et al. 2009, Cole et al. 2014).

In this section, I argue that an alternative view—the “mixture” or “superposition” view—can account equally well for the extant data. On that view, there is no straightforward way to



attribute functions to resting state networks, as functional connectivity “networks” may be a kind of *sampling artifact* rather than coherent functional units in the brain. Here is the general shape of the worry. Resting state fMRI involves measuring BOLD correlations over relatively long periods of time. Functional connectivity is typically calculated in terms of correlations between regions over an entire scanning session, which can last as long as multiple hours. Furthermore, scanning protocols for resting state fMRI sometimes involve longer acquisition times (TRs) than task-based fMRI, since the BOLD fluctuations measured at rest are slower than most task-related changes (see Power, Schlaggar, and Petersen 2014). The sum total of these different methodological choices is that resting state fMRI measurements indirectly capture the brain’s activity over a relatively long timescale.

As Morcom and Fletcher (2007) note, however, the “resting” brain is a complex, dynamic state, in which various functional regions and networks interact with one another as participants *imagine* what they are having for dinner, *remember* going to the dentist last week, *attend* to the noise of the scanner, and so forth. The extent to which these cognitive activities are taking place, and the extent to which regions or networks associated with them interact with one another, are unknown since the resting state is completely unconstrained from a cognitive standpoint. Therefore, there is no guarantee that the large-scale functional connectivity networks reflect coherent functional units; instead, they may reflect a mixture or superposition of smaller, task-related activations during the session. More generally, edges in the networks inferred by standard methods applied to resting state fMRI data are semantically ambiguous—they can indicate either actual connections or spurious correlations—and so we cannot necessarily infer the existence of a “default mode network,” even if multiple studies reveal similar networks.

It is important to be clear about the nature of the ‘mixtures’ being discussed here. At any given moment, the brain is a “mixture” (in the colloquial sense) of different functional processes; for example, reading involves word form recognition, attention, eye movement control, etc. all at once. Many researchers have dealt with the problem of how to disentangle these simultaneously occurring processes from one another. In fact, this is one of the issues that subtraction was intended to solve (Petersen and Fiez 1993). I have something different in mind: a mixture of processes occurring *over time* rather. That is, I suggest that resting state analyses involve mixtures of causally *and temporally* distinct brain processes: the mixture is of processes that occur at different times throughout the scanning session, and that potentially operate over different time scales. According to this proposal, a large-scale “network” identified in resting state research might never have been active as a single network during the scanning session; there might be no discrete time period during the scan in which a network with that topology was engaged. Instead, the “network” is simply an artifact of analyzing data from distinct sampling processes occurring at different times. I contend that the brain is a mixture in both senses—at a time and across time—but are here primarily focused on the problem of mixtures over time.

#### **4.3.2 Sampling a mixture distribution**

At a high level, a mixture distribution is a probability distribution that results from “mixing” distinct populations. One typically finds mixture distributions when multiple, heterogeneous populations are sampled without one knowing or measuring the populations from which each individual is drawn. For example, if one measured the hair length of randomly chosen students at a college, then the resulting distribution would likely be a mixture distribution, as the populations

of men and women (typically) have different distributions of hair length. There are many statistical questions and challenges relating to mixture distributions, such as how to estimate parameters of the mixed distributions, or how to determine whether a set of observations was probably drawn from a mixture distribution. Our primary interest here, though, is in the well-known fact that sampling from a mixture distribution can result in spurious correlations (e.g., Redner and Walker, 1984; Zheng and Frey 2004).

Consider the following example. Assume there is no correlation between height and beard growth in men; that is, a short man is just as likely to sport a beard as a tall one. Under these conditions, sampling from the population of men will likely yield a low correlation between height and beardedness. Now consider that, on average, women are both shorter than men, and much less likely to have a beard. If we take measurements of both men and women (and do not also measure gender), then we will likely find a correlation between height and beard growth. Qualitatively, the explanation for the correlation is that a short individual is more likely to be a woman and so less likely to have a beard, while a tall individual is more likely to be a man and so more likely to have a beard. That is, knowledge of an individual's height is informative about his or her beardedness, even though such knowledge is *uninformative* if one looks only at men, or only at women. In this toy example, the correlation between height and beardedness is spurious because (by assumption) they are uncorrelated in both of the populations. The correlation is a sampling artifact that results from drawing individuals from a mixture of two populations that are heterogeneous with respect to relevant distributional parameters.

This toy example involves a spurious correlation resulting from the mixture of populations in which the variables are uncorrelated. The opposite phenomenon can also occur:

two variables can be independent in the mixture distribution, even though they are correlated within each subpopulation. More generally, even if the qualitative existence and direction of correlations are the same in every subpopulation, the mixture distribution can have quite different correlations. Some classic statistical puzzles such as Simpson's Paradox and the Berkeley graduate admissions case of the 1970's (Bickel, Hammell, & O'Connell, 1975) arise because mixture distributions can have quite different qualitative features than any of the mixed subpopulations. The standard methodological advice in these cases is to analyze only statistically or causally homogeneous subpopulations, but this advice is of questionable use when our investigation is intended exactly to discover the causal relations that define these subpopulations (Cartwright, 1979, 1989).

I hypothesize that neuroimaging of the resting brain may actually involve sampling from a mixture distribution. As a result, the inter-region correlations in the data, and the connectivity networks that are subsequently inferred, must be interpreted with corresponding caution. As noted earlier, the term 'resting state' may be something of a misnomer since the "resting" participant is presumably continuing to reason, consider, imagine, visualize, and more while in the magnet. These different types of cognition are presumably heterogeneous with respect to neural firing; such differences are, after all, exactly what task-evoked fMRI aims to discover. Therefore, the "resting state" fMRI datapoints will be drawn from a mixture distribution, with the particular mixture being determined by the particular sequence of cognitions for the particular experimental participant. Moreover, to the extent that heterogeneous cognitions can "co-occur" (e.g., remembering playing basketball might involve both visual and motor areas), then the mixture will be both within- and across-times. Since the mixture distribution can exhibit quite different correlations than any of its component distributions, we cannot straightforwardly

interpret the output of network inference methods that depend on such correlations (or associations, more generally). Essentially all extant methods in neuroimaging, however, use association information as the basis for inference. In our toy example above, we “discovered” a connection between two variables where none exists in any individual or sub-population. The worry is that networks learned from resting state fMRI data might be similarly overinterpreted.

Although mixture distributions *can* be quite different than their components, certain properties are more plausible than others.<sup>39</sup> In particular, if two variables are correlated in many or all of the component distributions, then the mixture parameters typically must be set to precise values in order to yield independence (or zero correlation) in the mixture distribution. In contrast, two variables that are independent in every component distribution will typically be correlated in the mixture distribution for a wide range of mixture parameters. Put more colloquially, mixture distributions typically create correlations or associations, rather than eliminate them. The exception is when a variable has roughly the same probability distribution in each of the component distribution; in that case, mixing typically does not create new associations.

The mixture view thus leads to some qualitative expectations, even in the absence of knowledge of the particular mixture in a particular experimental participant. In general, we should not expect the networks inferred from resting state fMRI data to be complete (i.e., with an edge between every pair of variables), since some brain areas presumably have similar probabilities across different cognitions. However, we should expect to find that the inferred networks are supergraphs (i.e., have strictly more edges) of the superposition of the graphs for the (relevant) cognitions ([Figure 4.3](#)). Moreover, this is exactly the high-level pattern that has

---

<sup>39</sup> The claims that follow in this paragraph are all qualitative, but can be made precise by reading “implausible” as Lebesgue measure zero. I do not make stronger claims about probability or likelihood, however, as the standard Lebesgue measure might be inappropriate in some situations. I suspect that probability distributions over the parameters (of the mixture and component distributions) will likely exhibit significant context-dependence.

been found in many resting state fMRI papers: the inferred networks are supergraphs of various task-evoked networks that would plausibly be triggered during “free thought” (e.g., Cole, *et al.*, 2014). The mixture view implies that the edges in these networks (probably) come from two different sources—actual connections in one or more mixture component graphs, and spurious correlations resulting from the mixture itself—but the edges are not labeled in this way by (standard) network inference methods. We thus cannot reliably interpret any particular edge, subgraph, or even the whole inferred network. Perhaps it picks out a “default mode network” or perhaps specific edges correspond to actual neural connections, but existing methods do not tell us.

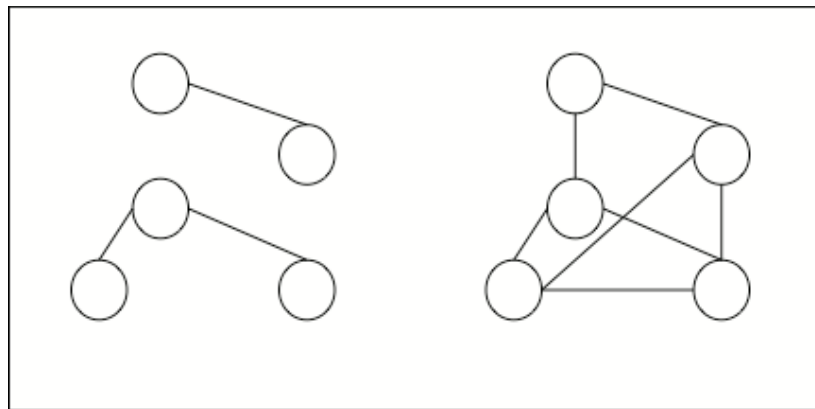


Figure 4.3: The mixture problem. Sampling from a mixture distribution can result in a structure that: (a) resembles the superposition of the constituent networks, and (b) has additional inferred edges.

### 4.3.3 More evidence for the mixture view

Mixture distributions present significant inferential challenges, and I have provided some reasons to think that resting state fMRI time series might constitute a mixture distribution. At the same time, these considerations show only that the mixture view is one explanation of the available data, not necessarily the best or only one. In this section, I consider additional evidence that favors the mixture view over competitors. If my arguments are right, then we should be agnostic (at least) about the leading interpretation of resting state analyses: namely, that they can be used to identify a small number of functionally relevant large-scale brain networks. Underdetermination about the source of edges in an inferred network complicates the attempt to give functional interpretations to stable functional connectivity patterns in resting state analyses—both *whether* Network X performs a specific function and *what* function Network X might perform. I emphasize, though, that the mixture view is *not* committed to the (overly strong) claim that every network in the mixture corresponds to some conscious cognition or task. The mixture view is entirely consistent with some networks, both within- and across-times, being responsible for unconscious cognition or purely physiological functions.

Allen, *et al.* (2012) give direct support for the mixture view, though on the analysis side rather than in data collection ([see Section 4.2](#)). They took resting state fMRI data on a large number of participants, but instead of finding a single network for the full time series, they inferred networks for much shorter segments of each individual time series.<sup>40</sup> They then used these networks to infer the “network timecourse” for each participant in their database, and

---

<sup>40</sup> Specifically, they first estimated the covariance matrices for moving windows of approximately 44 seconds, and then clustered those matrices. The centroid covariance matrices for those clusters were then used to infer connectivity networks.

showed that most participants had multiple networks over the course of data collection (typically changing every 5-10 seconds during the 5-minute scan). Moreover, this “changing networks” model fit the data significantly better than a “static network” model, even accounting for the additional degrees of freedom in the former model. Allen, *et al.* (2012) thus provide direct evidence that one of the fundamental presuppositions of the mixture view—namely, resting state fMRI time series data are generated by different networks over time—is plausibly correct. Interestingly, they found some “local” (in time) networks that are quite similar to the previously identified default mode network, which suggests that perhaps not all inferred edges are due solely to conscious, task-evoked neural activity.

Cole, *et al.* (2014) considered a question left open by Allen, *et al.* (2012): what is the relationship between the resting state network (inferred from the full time series) and various task-evoked networks? In general, Cole, *et al.* found that the resting state network was quite similar to the “multitask” network generated from multiple task-based (i.e., *not* resting state) studies. That is, the resting state network was similar to the superposition of networks produced by either 7 or 64 different tasks (e.g., working memory, motor, or decision making tasks). They additionally compared the resting state network against the network for each specific task, and found that the majority of the differences were edges present in the resting state network but absent in the task-evoked network. The mixture view predicts exactly this pattern of results since sampling from mixture distributions tends to generate correlations rather than eliminate them; the standard interpretation arguably does not. Interestingly, Cole, *et al.* (2014) do not endorse, or even seriously consider, the mixture view, but instead argue that the resting state network should be reinterpreted as describing the “fundamental architecture” of the brain: that is, it tells us the



possible connections between regions, rather than the connections used in performing any particular cognitive task.

Andrews-Hanna, Smallwood, and Spreng (2014) provide evidence that further supports the mixture view in their investigation of potential subsystems composing the default mode network. Yeo and colleagues (2011) used a clustering algorithm to identify three default mode network subsystems based on functional connectivity: a “core” network, a dorsal medial subsystem, and a medial temporal subsystem. Andrews-Hanna, Smallwood, and Spreng performed a meta-analysis in NeuroSynth,<sup>41</sup> asking what functional labels (from task-based studies) are associated with each default mode network subsystem identified by Yeo and colleagues. Interestingly, there were consistent functional differences between these three subsystems. For example, the dorsal medial subsystem was strongly associated with semantics and theory of mind, while the medial temporal subsystem was associated with episodic memory. These findings suggest that perhaps the default mode network is not a single functional network, but a collection of associated networks supporting different cognitive functions.

#### **4.4 OBJECTIONS AND REPLIES**

If the mixture view is correct, then resting state fMRI research will need to adopt new protocols and analyses for identifying functional brain networks in order to account for mixtures ([see Section 4.5](#)). Before turning to those positive recommendations, though, I consider three main objections that proponents of resting state fMRI have raised, or could raise, against the mixture view. First, the timing and magnitude of BOLD fluctuations observed in resting state analyses do

---

<sup>41</sup> A database of fMRI images, optimized for meta-analyses: <http://www.neurosynth.org>

not seem to resemble task-evoked BOLD fluctuations. Second, resting state functional connectivity patterns exhibit more consistency (within- and between-subjects) than the mixture view would predict. Third, resting state activity persists in the absence of conscious cognition (e.g., in anesthetized patients). In each case, I argue that the objection fails to rule out the mixture or superposition interpretation of resting state functional connectivity patterns.

#### **4.4.1 Objection 1: consistency of resting state functional connectivity patterns**

Many studies report that resting state functional connectivity patterns are highly consistent both between and within subjects (Damoiseaux et al. 2006, Laumann et al. 2015). Researchers have found a substantial degree of between-subject consistency: roughly similar functional connectivity patterns emerge in different subjects (Damoiseaux et al. 2006) and even in non-human mammals such as mice (White et al. 2011) or monkeys (Vincent et al. 2007). They have also found within-subject consistency: idiosyncratic features of functional connectivity patterns, as well as the magnitude of resting BOLD fluctuations, are consistent across different scanning sessions for the same participant (Damoiseaux et al. 2006, Laumann et al. 2015). One might argue that the mixture view, which holds that functional connectivity patterns are mixtures of different functional networks, would not predict a high degree of within-subject or between-subject consistency. The following example will illustrate the concern.

One day, a participant comes into the scanner. She thinks about an upcoming trip to Disney World, vividly imagining the food (which evokes activity in gustatory cortex), riding the roller coasters (which evokes activity in visual and vestibular areas), and so forth. The next scanning session, the same participant spends the entire time playing songs in her head from a

concert she attended that week (which evokes activity in the auditory cortex). If resting state functional connectivity patterns are mixtures of conscious cognitive processes, we might expect quite different functional connectivity patterns to emerge from these sessions. Simply put, the worry is that since different time blocks of conscious cognition involve a mixture of different cognitive processes, the functional connectivity patterns arising from different scanning sessions should not be particularly consistent, either within- or between-subjects.

This objection presupposes, however, that there is little-to-no consistency in the brain networks activated during the “free range” cognition that occurs in a resting state experiment. Different people and possibly other mammals, might be engaged in a common set of background functions during the scanning session—the proprioceptive system is monitoring changes in limb position, the oculomotor system is controlling eye movements, the auditory system is responding to scanner noises, etc. There also are plausibly a number of “typical” tasks, perhaps quite basic ones, which recruit a host of cognitive functions that follow one another in a predictable fashion. For example, accessing an episodic memory (e.g., remembering a trip to the dentist yesterday) might often be accompanied by mental imagery (e.g., imagining how one’s lips felt after the novacaine injection) and often succeeded by planning (e.g., thinking about a doctor’s appointment next week). Thus, a brain engaged in “free range” cognition might exhibit a relatively stable distribution of ongoing processes, as well as patterns in the order and temporal proximity of recruited processes, and other network activations.

Moreover, the mixture view predicts that the large-scale functional connectivity networks identified in resting state analyses are often mixtures of *related* cognitive processes, since processes that co-occur, or succeed one another regularly, would be mixed together more often. Provocatively, the many different functions Andrews-Hanna, Smallwood, and Spreng (2014)

associate with the default mode network (e.g., social cognition, introspection, autobiographical memory, etc.) seem to be connected in just this way. It is plausible that social interactions frequently involve the internal monitoring of emotional or belief states, that episodic memory is often accompanied by introspection, etc. Whether or not this story holds, or if a similar story can be told for other ICNs (e.g., large-scale networks associated with attention, motor control, etc.), remains to be seen. Our main point here is that engaging in free range cognition over the course of a long scan session can result in between subject or within subject consistency in functional connectivity patterns, so long as cognitive processes occur, co-occur, and follow one another, in a regular fashion.

#### **4.4.2 Objection 2: timing and magnitude of resting state fluctuations**

Some authors report that the timing and magnitude of resting state BOLD fluctuations are different than task-evoked BOLD fluctuations (Raichle 2009, Snyder and Raichle 2012). Snyder and Raichle (2012) argue that resting state BOLD fluctuations are both slower and larger in magnitude than task-evoked BOLD changes. For them, these quantitative differences suggest that, “unconstrained cognition alone does not account for the greatest part of intrinsic activity” (Snyder and Raichle 2012, 903). However, the mixture view can easily account for these qualitative differences in timing between task-evoked and resting state BOLD fluctuations: if functional connectivity patterns arise from mixtures of cognitive processes operating at different time scales, then these mixtures will take place over longer time scales than their component processes (i.e., the phenomena that neuroscientists typically measure using cognitive tasks). As for the magnitude of BOLD signal changes, there are conflicting results. While Snyder and

Raichle (2012) report that resting BOLD fluctuations are greater in magnitude than task-evoked fluctuations, Damoiseaux and colleagues (2006) report that resting state fluctuations are typically the same magnitude as task-evoked activations. Whether this discrepancy is due to some feature of the experimental protocol (e.g., different instructions of what to do during “rest”) or the fact that the resting state is a mixture of both free range cognition and intrinsic signal remains to be seen. In fact, I suggest that this disagreement points towards a possible way to test some implications of the mixture view.

#### **4.4.3 Objection 3: resting state fluctuations persist without consciousness**

A more serious objection to the mixture view comes from the finding that resting state activity persists in heavily sedated monkeys (e.g., Vincent et al. 2007) and humans (e.g., Greicius et al. 2008). To the extent that the mixture is composed of conscious cognitions, this result is quite surprising. Functional connectivity patterns measured during rest in alert participants seem to persist even in unconscious individuals, which makes it more plausible that resting state fMRI is measuring intrinsic activity.

There are two points to make here. First, as I noted earlier, the mixture view is not committed to the idea that *all* resting state functional connectivity patterns arise from conscious cognition. The brain undoubtedly exhibits some degree of baseline metabolic activity, and essentially all cognitive theories leave room for ongoing or background activity. For example, predictive coding theories of the brain (e.g., Hohwy 2014) hold that the brain is constantly generating predictive models of the environment, and testing these models against incoming information, whether from external sensory input or other sources (e.g., other parts of the brain).

The mixture view is entirely consistent with functional connectivity patterns in resting, alert participants being a mix of baseline metabolic processes (neural activity present even in semiconscious patients), background cognitive processes (e.g., predictive coding), and free-ranging conscious cognition. On the mixture view, the stability of certain resting state networks in sedated individuals (e.g., Rosazza and Minati, 2011) provides insight into one of these process-types, rather than implying that the mixture view is false. Moreover, studying some of these processes (e.g., conscious or background cognition) may require demixing methods of exactly the sort that I suggest in [4.5](#). For instance, using resting state fMRI to identify novel networks involved in conscious cognition might require demixing conscious activity from baseline metabolic activity.

Second, the problem of mixtures I am identifying is not *just* a problem for studying conscious cognition, but a sampling problem that could also apply to purely unconscious cognitive or metabolic processes. Even if resting state protocols measure intrinsic neural activity, this activity could result from many distinct processes that are mixed together in the sampling process. Some proponents of resting state fMRI argue that resting state analyses measure intrinsic activity—i.e. activity that would persist in the absence of conscious cognition—but that this activity reflects preparatory or anticipatory activity in known functional systems—e.g., motor or visual areas (Damoiseaux et al. 2006). Imagine that one could tell from an idling car engine what parts work together when the engine is moving; this is roughly akin to identifying what brain regions work together in the resting (think idling) brain. But for resting state fMRI to play the role of identifying functional subsystems, whether or not the activity in these systems results from free-ranging cognition or baseline metabolic activity, researchers will need to account for the possibility of mixtures.

## **4.5 THEORETICAL AND METHODOLOGICAL CONSEQUENCES**

I now consider some of the methodological and theoretical implications of the mixture view, as it implies that resting state data should be collected, analyzed, and interpreted differently, though there are significant challenges with all of these changes.

### **4.5.1 New tools and techniques**

The most obvious change implied by the mixture view is the need to use various mathematical and statistical techniques to try to disentangle the component networks from which the mixture is constructed. This is a particularly challenging problem, as there can be mixtures of networks both within and across times. Moreover, the component networks are not necessarily known in advance; determining the network structure for various cognitive activities is itself a major research challenge. We thus cannot simply look for evidence of different known networks, but rather must simultaneously infer the networks and their mixing parameters.

There are well-established techniques for solving parts of the mixing problem, but no methods that can simultaneously handle all of the different complexities. For example, standard techniques such as independent components analysis (ICA) can extract multiple networks that jointly produce the observed time series data. As a result, ICA is sometimes thought to be a solution to mixture worries about resting state data. These methods are only reliable, however, if the mixture is relatively stable during the relevant time period, which is exactly what I suggest does not hold. Alternately, there are a number of changepoint detection algorithms (Adams and

MacKay, 2007; Desobry, *et al.*, 2005) or other methods for time series segmentation (e.g., Gregory, Nason, and Watt, 1996; Scargle, 1998) that can determine time points at which it is likely that the underlying generating system has changed. One could then presumably use methods such as ICA within each time period to extract the component networks, though the reliability of this combined method would need to be evaluated. An additional complexity is that standard changepoint detection methods assume that there are relatively sharp breaks or discontinuities in the evolution of the generating system, while the mixture view allows for the possibility that the mixture changes slowly over time. Overall, it seems plausible that tools could be developed to “demix” resting state time series, but reliable, “end-to-end” methods have not yet been developed.<sup>42</sup>

Even given such methods, however, many resting state experimental protocols arguably do not collect sufficiently fine-grained data to adequately separate the mixture components or changing mixture parameters.<sup>43</sup> Most resting state researchers have used longer repetition times (TRs) in their studies, as these magnet settings yield improved signal-to-noise ratios without impairing inference about the slower changes over longer timescales that have been the principal focus of resting state research. More recently, resting time studies have shifted towards using TRs of approximately 2000 ms. The mixture view argues, however, that significant components of the “resting” state time series are due to networks for cognitive activities that are presumably changing even more rapidly than this. Thus, these data arguably lack the temporal resolution to adequately separate the mixture components (though see, e.g., Allen, *et al.*, 2012). It is an open

---

<sup>42</sup> Of course, this optimism ignores the problem of inferring networks from correlations given complications such as time-averaging of neural activity in the BOLD signal, undersampling of fMRI measurements, and many other issues (e.g., Seth, Chorley, & Barnett, 2013).

<sup>43</sup> Thanks to David Plaut for emphasizing this worry.



question whether typical fMRI measurements are sufficiently “clean” and fast to enable reliable learning of network mixture components.

A different, complementary approach to the mixture problem would be to collect additional data that further constrains the possibility space for the mixture parameters. In particular, the mixture view holds that at least some of the component networks are responsible for the individual’s conscious cognition; they are the networks responsible for *remembering* the grocery list, *planning* one’s route from the lab to the store, *deciding* whether to stop on the way home, and so on. Resting state experiments do not normally collect reports about participants’ conscious cognition or introspective mental activities. In principle, information about participants’ internal cognition can provide valuable constraints on the identification of mixture components and parameters. If an individual is visualizing at particular times, for example, then there should be certain similarities in the inferred networks for those times. Self-reports can potentially provide key information for demixing methods, even granting that conscious, introspectively accessible cognition is informative about only some of the brain’s activity.

There are, however, serious concerns about obtaining these types of cognitive constraints. First, there are long-standing worries about the reliability of introspection and self-reporting (Jones and Harris, 1967; Nisbett and Ross, 1991; Schwitzgebel, 2008; Engelbert and Carruthers, 2010). Some of these worries are not necessarily an issue in this context; participants would not be asked to report, for example, their reasons for action. Not all such concerns can be so easily dismissed, however, and so it is unclear whether these cognitive data would be sufficiently reliable to provide useful constraints. Second, there are methodological challenges to obtaining these reports. If participants are asked to report their conscious cognition in real-time, then the study will involve a constant task (namely, to remember one’s conscious states) and so no longer

be about the resting state. Even if participants give reports only afterwards, knowledge that they will be asked to do so can plausibly induce constant memory and metacognition tasks throughout the experiment. Thus, it seems that the request for (retrospective) self-reports must come as a surprise at the end of the study. In that case, though, one worries that participants will be relying on whatever information was ad hoc encoded, perhaps with significant error, in memory. It is simply an open empirical question whether those self-reports would be more unreliable. In any case, there are clearly substantial experimental challenges to obtaining self-reports of conscious cognition, even though such data could significantly improve demixing analyses.

#### **4.5.2 The value of resting state data**

An alternate, though not mutually exclusive, response is to see a different kind of value in resting state data. The standard framing of discussion of these experiments has been based on a distinction between intrinsic and task-evoked networks: resting state studies inform us about the former; task-specific studies inform us about the latter. The mixture view challenges, however, the very idea that there is a stable intrinsic network being learned from resting state data. Perhaps there is some persistent background network in the brain, but if the mixture view is correct, then it cannot be inferred transparently from resting state fMRI data. Thus, different experiment-types cannot be distinguished by the types of functional networks they target.

As I suggested earlier, the core difference between these two types of experiments seems to be in the degree of experimental control (over brain activity) exercised by the researcher. Standard task-based studies involve significant control, as the participants' cognition is presumably driven principally by experimental demands. In contrast, resting state studies are

relatively uncontrolled, often quite explicitly so. A common view is that greater experimental control is always better, but this claim is not correct in general; rather, it depends on what one is trying to learn. Significant control (including, e.g., randomization) can be the best way to discover whether some particular *A* causes some particular *B*. If one is instead trying to understand, say, the typical behavior of a system in its natural environment, then significant experimental control can actually impair learning, precisely because that control can change the system from its natural state.

In many cases, scientists rely on exploratory experiments to map how a system operates in a “natural” setting—e.g., microarray studies can measure how gene transcription in a cell changes as the cell grows and divides (Franklin 2005). The mixture view implies that resting state fMRI studies could be similarly valuable for answering questions about “natural” brain behavior. Task-based studies are, by design, artificial in certain respects, and so potentially misleading about aspects of neural activity in more naturalistic settings. In particular, task-based studies push the individual to engage in certain cognitive activities, typically to the exclusion of others. As a result, there are potentially significant brain networks that have not been observed, simply because those networks underlie some task that has not been isolated in any particular experiment. Precisely because resting state studies are relatively uncontrolled, they hold forth the promise of revealing previously unknown or understudied brain networks. People’s free-ranging cognition plausibly traverses a wider space than experimenters have previously thought to isolate.

I want to emphasize that this conclusion is decidedly *not* a skeptical one. I am not arguing that the mixture view implies that all inferences from resting state connectivity patterns to functional networks must be ambiguous, or that all resting state studies are worthless. I

vehemently disagree with this assessment. My point is rather that we must be careful about the questions that we expect resting state studies to answer. They can be incredibly useful and powerful, but they must be interpreted with appropriate care. On the mixture view, these studies do not necessarily reveal some intrinsic, task-free, omnipresent network. They can, however, be used to discover new brain networks and better understand naturalistic brain activity, including the ways in which different networks co-occur and interact when external conditions do not dictate a particular goal, task, or cognition. That is, resting state fMRI studies can be (I suggest) a full-blooded realization of the possibilities of exploratory science. In order to do so, however, we must ensure that we account for mixtures by adopting the methodological and analytic techniques outlined earlier.

## **4.6 EXPLORING BRAIN NETWORKS**

Recent work in the philosophy of experiment suggests that exploratory experiments can reveal interesting patterns about a system in the absence of explicit theories of how those patterns arise (Steinle 1997, Hall 2005). Thus successful experiments need not involve testing specific hypotheses under controlled settings. Resting state fMRI fits this description: it putatively identifies functional brain networks (via functional connectivity patterns) in a more naturalistic setting. However, as Franklin-Hall (née Franklin) notes (Franklin 2005), exploratory experiments rarely meet the Baconian ideal of theory-free observation. Instead, interpreting exploratory experiments involves a dense constellation of theoretical commitments about both instrumentation and the workings of the system in question.

In this chapter, I analyzed the theoretical commitments of resting state fMRI. Many researchers argue that the principal difference between task-based fMRI and resting state fMRI is that the former measures activity related to stimulus or goal driven cognitive functions while the latter measures intrinsic or internally driven brain activity. Researchers interpret the main finding in resting state research—i.e. the presence of consistent, large-scale functional connectivity patterns—in a number of different ways. However, a common theme in interpreting these results is that resting state analyses can be used to identify large-scale functional networks in advance of knowing what exactly those networks are doing. There is much more agreement on this point than on what networks there are, or what precise functions these networks may be performing.

My account challenges this central way of interpreting resting state research. The mixture view proposes that the large-scale networks identified in resting state fMRI may be sampling artifacts rather than genuine features of the brain's functional organization. According to this view, the brain exhibits different kinds of processes (background metabolic, background cognitive, and free-ranging conscious cognition) that occur at different times and operate over different time scales. Since it is likely that resting state analyses involve mixing these temporally and causally distinct processes, there is currently no way to be sure that the topological “structures” identified in resting state research correspond to functional brain networks. Nevertheless, we agree that resting state analyses hold the promise of identifying novel functional networks, since free-ranging cognition can traverse a broader range of brain processes than researchers typically target in task-based studies. Accounting for the possibility of mixtures will improve the ability of resting state analyses to disentangle these diverse functional signals, and thus make good on the promise of identifying candidate functional networks in a bottom-up fashion.

## 5.0 DISSERTATION CONCLUSION

### 5.1 THE STORY SO FAR

In this dissertation, I have demonstrated that fMRI studies can, in principle, inform the development and assessment of cognitive theories. I have also shown that the value of fMRI for testing cognitive theories depends on the *bridging assumptions* that neuroimagers use to link BOLD activation patterns to cognitive theories. The main philosophical problem is that recent developments in brain mapping—both theoretical developments such as context-sensitivity (McIntosh 2000) and network-oriented mapping (Sporns 2011), and methodological developments such as MVPA (Haxby et al. 2001) and resting state fMRI (Snyder and Raichle 2012)—cast serious doubt on the assumptions that have typically made fMRI results speak to cognitive theorizing.

I conclude that fMRI should be used not just to test cognitive theories (indeed, given the discussion in [Chapter 3](#) it may be premature to expect fMRI to be useful for validating psychological constructs), but must also be used to test the bridging assumptions on which cognitive inferences rely. This is an extension of the kind of project Poldrack (2006) and Anderson (2010) undertook when they questioned the extent to which individual regions are recruited for different cognitive capacities, and what this means for reverse inference in cognitive neuroimaging. I recognize, however, that this general picture raises far more questions than it

answers. In what follows, I briefly outline some outstanding questions, and how one might approach them.

## 5.2 BRAIN MAPPING: HOW BAD COULD IT BE?

[In Chapter 2](#), I presented a “modest” picture of the revisions neuroscientists will need to make when mapping functions onto the brain. Minimally, we should expect individual regions to be frequently re-deployed (Anderson 2010) and that many psychological capacities will map onto networks rather than any particular region (Sporns 2011). This picture has many consequences—for one, it predicts that selective associations between structures and functions may emerge only at the level of large-scale brain networks. But there is no *guarantee* that psychological kinds will map neatly onto sets of brain regions. First, functional brain networks are not just collections of regions. The same regions may perform different functions when *connected* differently. For example, the same three regions may predict what perceptual inputs will result from performing a motion when engaged one way, and select which motion to perform based on perceptual inputs when engaged another way. This is why studying changes in effective connectivity, which (unlike the analyses I have examined so far) reveal the direction of influence between regions, is so important (Friston 2011, Pessoa 2014).

Even more vexing, it is possible that the problem of multi-functionality I discussed in [Chapter 2](#)—i.e. that the same parts can perform different operations in different contexts—will recur at larger *scales* within the brain. According to this concern, large-scale brain networks may perform different operations depending on what coalitions of large-scale networks they are currently interacting with, just as regions seem to perform different functions in different neural

contexts. This kind of picture would greatly complicate inferences from fMRI results to cognitive theories. Thus, the same basic conceptual problems may recur as neuroscientists adopt new tools—e.g., resting state fMRI—for observing brain activity at different scales. Of course, determining whether and in what sense a large-scale brain network is multi-functional will require dealing with problems like the mixture problem discussed in [Chapter 4](#), since the appearance of a large, multi-functional network may be a sampling artifact in some instances.

Another issue concerns the extent to which analyses developed to understand one brain system will apply to another. My account in [Chapter 2](#) suggests that not only will each brain region perform some characteristic operation, but also each region may be capable of performing different operations in different contexts. While it remains speculative that this more radical form of context-sensitivity (i.e. performing different computations or operations in different contexts) is found within human cortex, it is found in a number of other biological systems. But while I emphasize functional heterogeneity, a number of neuroscientists (e.g., Carandini and Heeger 2012) and philosophers (e.g., Chirimuuta 2014) have argued that neural circuits reuse the same canonical computations in a number of different capacities. If this kind of picture is true, then perhaps neuroscientists will be able to understand cognition as the combination and recombination of a fairly circumscribed number of computational elements. The extent to which different regions perform the same kinds of operations is an important, yet unresolved, theoretical issue.



### 5.3 COGNITIVE ONTOLOGY REVISION: A ROADMAP

My dissertation raises a number of questions about what fMRI can contribute to testing our taxonomy of psychological kinds. Poldrack (2010) advocates testing our cognitive ontologies in a “top down” fashion. This project involves carefully spelling out the taxonomic relationships between our cognitive constructs—e.g., response inhibition, task switching, and working memory—and then using fMRI to test whether or not these constructs seem to correspond to the same brain processes (Lenartowicz et al. 2010, Poldrack 2010). In this approach, scientists can develop their cognitive ontology independently of brain data, and then use brain data to validate and refine it. By contrast, Anderson (2014, Ch. 4) articulates a radical, “bottom up,” form of fMRI-based cognitive ontology revision. In this approach, neuroimagers will use dimension reduction analyses to uncover common activation patterns between disparate tasks, and thereby identify novel cognitive constructs.

But perhaps both strategies are the wrong way to go about revising our cognitive ontologies. Instead of framing their experiments as a “test” of some cognitive theory, many neuroscientists are instead merely attempting to redefine the functions of brain areas. For example, Aminoff, Kveraga, and Barr (2013) developed the notion of “contextual processing,” to explain the parahippocampal cortex’ involvement in spatial navigation, scene processing, and a host of other functions. While it is hard to find examples where fMRI data outright contradicts some cognitive theory ([see Chapter 3](#)), neuroscientists seem to be developing new vocabularies for describing the contribution of brain regions to cognition. Gauthier et al. (2001) talk of “visual expertise” rather than face recognition. Menon and Uddin (2010) implicate the anterior insula in “salience detection,” a form of metamodal processing involved in detecting both internal and

external changes in the environment. Perhaps this process of re-describing the functions of brain regions (and developing computational models based on these reconceived functions) will result in a kind of cognitive ontology revision that little resembles using fMRI to “test” our cognitive theories.

## 5.4 MENTAL FUNCTION AND CEREBRAL CARTOGRAPHY

The central lessons of this dissertation are as follows. First, theories of the brain’s functional topography *interact* with, and in many cases determine, the value of fMRI for testing cognitive theories. The taxonomic and cartographic projects outlined in [Chapter 1](#) are deeply dependent on each other. Second, since theories of the brain’s functional topography are currently in flux, so too are the cognitive inferences neuroscientists can draw from fMRI data. Therefore, neuroscientists will need to use fMRI, and other methods, to test these bridging assumptions. Finally, the bridging assumptions that ultimately bring fMRI into contact with cognitive theories will likely need to be far more “local”—i.e. specific to inferences in particular experimental paradigms, or about particular brain systems—than previous researchers have recognized.

## BIBLIOGRAPHY

- Adams, R. P., and MacKay, D. J. C. 2007. "Bayesian Online Changepoint Detection." Technical Report, University of Cambridge, Cambridge, UK. arXiv:0710.3742v1 [stat.ML].
- Aguirre, G. K. 2014. "Functional Neuroimaging: Technical, Logical, and Social Perspectives." *Hastings Center Report*, 44(s2): S8-S18.
- Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., & Calhoun, V. D. 2012. "Tracking whole-brain connectivity dynamics in the resting state." *Cerebral Cortex*, 24 (3): 663-676
- Aminoff, E.A., Kveraga, K., and Bar, M. 2013. "The Role of Parahippocampal Cortex in Cognition." *Trends in Cognitive Sciences*, 17: 379-390.
- Amundson, R. and Lauder, G.V. 1994. "Function without Purpose: The Uses of Causal Role Function in Evolutionary Biology." *Biology and Philosophy*, 9: 443-469.
- Anderson, M. L. 2010. "Neural Reuse: A Fundamental Organizational Principle of the Brain." *Behavioral and Brain Sciences*, 33: 245-313.
- Anderson, M.L. 2014. *After Phrenology: Neural Reuse and the Interactive Brain*. Cambridge, Massachusetts: MIT Press.
- Anderson, M. L. 2015. "Mining the Brain for a New Taxonomy of the Mind." *Philosophy Compass*, 10(1): 68-77.
- Andrews-Hanna, J. R., Smallwood, J., & Spreng, R. N. 2014. "The Default Network and Self-Generated Thought: Component Processes, Dynamic Control, and Clinical relevance." *Annals of the New York Academy of Sciences*, 1316(1): 29-52.
- Arbabshirani, M. R., Castro, E., & Calhoun, V. D. 2014. "Accurate Classification of Schizophrenia Patients based on Novel Resting-state fMRI Features." In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE* (pp. 6691-6694). IEEE.
- Banich, M.T. and Compton, R. 2010. *Cognitive Neuroscience*. Nelson Education.

- Baddeley, A., 2000. "The Episodic Buffer: a New Component of Working Memory?." *Trends in Cognitive Sciences*, 4(11): 417-423.
- Baddely, A. 2003. "Working Memory: Looking Back and Looking Forward." *Nature Reviews Neuroscience*, 4: 829-839.
- Baddeley, A.D. and Hitch, G., 1974. "Working memory." *The Psychology of Learning and Motivation*, 8: 47-89.
- Basso, K., Margolin, A. A., Stolovitzky, G., Klein, U., Dalla-Favera, R., & Califano, A. 2005. "Reverse Rngineering of Regulatory Networks in Human B Cells." *Nature Genetics*, 37(4): 382-390.
- Bechtel, W., 2005. "The Challenge of Characterizing Operations in the Mechanisms Underlying Behavior." *Journal of the Experimental Analysis of Behavior*, 84(3), pp.313-325.
- Bechtel, W., 2008. *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. Taylor & Francis.
- Bechtel, W. and Richardson, R. 1993. *Discovering Complexity*. Princeton, New Jersey: Princeton University Press.
- Beckmann, C. F., DeLuca, M., Devlin, J. T. & Smith, S. M. 2005. "Investigations into Resting-state Connectivity using Independent Component Analysis." *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 360: 1001-1013.
- Bergeron, V. 2007. "Anatomical and Functional Modularity in Cognitive Science: Shifting the Focus." *Philosophical Psychology*, 20 (2): 175-195.
- Bickel, P. J., Hammel, E. A., & O'Connell, J. W. 1975. "Sex Bias in Graduate Admissions: Data from Berkeley." *Science*, 187 (4175): 398-404.
- Bickle, J., 1995. "Psychoneural Reduction of the Genuinely Cognitive: Some Accomplished Facts." *Philosophical Psychology*, 8(3): 265-285.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Rao, S. M., & Cox, R. W. 1999. "Conceptual Processing During the Conscious Resting State: A functional MRI Study." *Journal of Cognitive Neuroscience*, 11: 80-95.
- Biswal, B. B., Yetkin, F. Z., Haughton, V. M., & Hyde, J. S. 1995. "Functional Connectivity in the Motor Cortex of Resting Human Brain using Echo-planar MRI." *Magnetic Resonance Imaging*, 34: 537-541.
- Biswal, B. B., Mennes, M., Zuo, X. N., Gohel, S. Kelly, C., Smith, S. M., Beckmann, C. F., Adelstein, J. S., Buckner, R. L., Colcombe, S., Dogonowski, A. M., Ernst, M., Fair, D.,

- Hampson, M., Hoptman, M. J., Hyde, J. S., Kiviniemi, V. J., Kotter, R., Li, S. J., Lin, C. P., Lowe, M. J., Mackay, C., Madden, D. J., Madsen, K. H., Margulies, D. S., Mayberg, H. S., McMahon, K., Monk, C. S., Mostofsky, S. H., Nagel, B. J., Pekar, J. J., Peltier, S. J., Petersen, S. E., Riedl, V., Rombouts, S. A., Ryma, B., Schlagger, B. L., Schmidt, S., Seidler, R. D., Siegle, G. J., Sorg, C., Teng, G. J., Veijola, J., Villringer, A., Walter, M., Wang, L., Weng, X. C., Whitfield-Gabrieli, S., Williamson, P., Windischberger, C., Zang, Y. F., Zhang, H. Y., Castellanos, F. X., & Milham, M. P. 2010. "Toward Discovery Science of Human Brain Function." *Proceedings of the National Academy of Sciences*, 107: 4734-4739.
- Bray, S., Arnold, A. E., Levy, R. M., & Iaria, G. 2015. "Spatial and Temporal Functional Connectivity Changes Between Resting and Attentive States." *Human Brain Mapping*, 36(2): 549-565.
- Brigman, K. L. and Kristan, W. B. 2008. "Multifunctional Pattern-Generating Circuits." *Annual Review of Neuroscience*, 31: 271-294.
- Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. 2008. "The Brain's Default Network: Anatomy, Function, and Relevance to Disease." *Annals of the New York Academy of Sciences*, 1124: 1-38.
- Bueti, D. and Walsh, V. 2009. "The Parietal Cortex and the Representation of Time, Space, Number, and other Magnitudes." *Philosophical Transactions of the Royal Society of London B*, 364(1525): 1831-1840.
- Calder, A.J., Keane, J., Manes, F., Antoun, N. and Young, A.W., 2000. "Impaired Recognition and Experience of Disgust Following Brain Injury." *Nature Neuroscience*, 3(11): 1077-1078.
- Campbell, D.T. and Fiske, D.W. 1959. "Convergent and Discriminant Validation by the Multitrait-multimethod Matrix." *Psychological Bulletin*, 56(2): 81.
- Campbell, D.T., 1960. "Recommendations for APA Test Standards Regarding Construct, Trait, or Discriminant Validity." *American Psychologist*, 15(8): 546.
- Carandini, M. and Heeger, D.J., 2012. "Normalization as a Canonical Neural Computation." *Nature Reviews Neuroscience*, 13(1): 51-62.
- Cartwright, N. 1979. "Causal laws and effective strategies." *Noûs*, 13 (4): 419-437.
- Cartwright, N. 1989. "*Nature's Capacities and their Measurement*." Oxford: Oxford University Press.
- Charney, D.S. and Deutch, A., 1996. "A Functional Neuroanatomy of Anxiety and Fear: Implications for the Pathophysiology and Treatment of Anxiety Disorders." *Critical Reviews in Neurobiology*, 10(3-4).

- Chatham, C. H., and Badre, D. (2015). "How to Test Cognitive Theory with fMRI." In D. Spieler & E. Schumacher (Eds.), *New Methods in Cognitive Psychology*. New York, NY: Routledge (in press).
- Chirimuuta, M., 2014. "Minimal Models and Canonical Neural Computations: The Distinctness of Computational Explanation in Neuroscience." *Synthese*, 191(2): 127-153.
- Church, J. A., Fair, D. A., Dosenbach, N. U., Cohen, A. L., Miezin, F. M., Petersen, S. E., & Schlaggar, B. L. 2009. "Control Networks in Paediatric Tourette Syndrome Show Immature and Anomalous Patterns of Functional Connectivity." *Brain*, 132(1): 225-238.
- Churchland, P.M., 1981. "Eliminative Materialism and the Propositional Attitudes." *The Journal of Philosophy*, 78(2): 67-90.
- Churchland, P.S., 1986. *Neurophilosophy: Toward a Unified Understanding of the Mind-Brain*. The MIT Press.
- Cohen, N.J. and Squire, L.R., 1980. "Preserved Learning and Retention of Pattern-Analyzing Skill in Amnesia: Dissociation of Knowing How and Knowing That." *Science*, 210(4466): 207-210.
- Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S., & Petersen, S. E. 2014. "Intrinsic and Task-Evoked Network Architectures of the Human Brain." *Neuron*, 83: 238-251.
- Coltheart, M. 2004. "Brain Imaging, Connectionism, and Cognitive Neuropsychology." *Cognitive Neuropsychology*, 21(1): 21-25.
- Coltheart, M. 2006. "Perhaps Functional Neuroimaging Has Not Told Us Anything About The Mind (So Far)." *Cortex*, 42(3): 422-427.
- Coltheart, M. 2013. "How Can Functional Neuroimaging Inform Cognitive Theories?" *Perspectives on Psychological Science*, 8(1): 98-103.
- Cooper, R.P. and Shallice, T., 2010. "Cognitive neuroscience: The Troubled Marriage of Cognitive Science and Neuroscience." *Topics in Cognitive Science*, 2(3): 98-406.
- Corkin, S., 1968. "Acquisition of motor skill after bilateral medial temporal-lobe excision." *Neuropsychologia*, 6(3): 255-265.
- Craver, C. F. 2001. "Role Functions, Mechanisms, and Hierarchy." *Philosophy of Science*, 68: 53-74.
- Craver, Carl F. 2007. *Explaining the Brain*. Oxford University Press.

- Craver, C.F., 2009. "Mechanisms and Natural Kinds." *Philosophical Psychology*, 22(5): 575-594.
- Cronbach, L.J. and Meehl, P.E. 1959. "Construct validity in psychological tests." *Psychological Bulletin* 52 (4): 281.
- Cummins, R. 1975. "Functional Analysis." *Journal of Philosophy*, 72: 741-765.
- Dailey, F.E. and Cronan, J.E. Jr. 1986. "Acetohydroxy Acid Sythase I, a Required Enzyme for Isoleucine and Valine Biosynthesis in *Escherichia coli* K12 During Growth on Acetate as the Sole Carbon Source." *Journal of Bacteriology*, 165(2): 453-460.
- Damarla, S.R., Cherkassky, V.L. and Just, M.A., 2016. "Modality-independent Representations of Small Quantities based on Brain Activation Patterns." *Human Brain Mapping*, 37(4): 1296-1307.
- Damoiseaux, J. S., Rombouts, S. A. R. B., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., & Beckmann, C. F. 2006. "Consistent Resting-state Networks Across Healthy Subjects." *Proceedings of the National Academy of Sciences*, 103(37): 13848-13853.
- Danks, D., 2015. "Goal-dependence in (Scientific) Ontology." *Synthese*, 192(11): 3601-3616.
- Darwin, C., Ekman, P. and Prodger, P., 1998. *The Expression of the Emotions in Man and Animals*. Oxford University Press, USA.
- Dawkins, M.J.R. 1966. "Biochemical Aspects of Developing Function in New-Born Mammalian Liver." *British Medical Bulletin*, 22(1): 27-33.
- De Brigard, F., Addis, D.R., Ford, J.H., Schacter, D.L. and Giovanello, K.S., 2013. "Remembering What Could have Happened: Neural Correlates of Episodic Counterfactual Thinking." *Neuropsychologia*, 51(12): 2401-2414.
- Dehaene, S., Cohen, L., Sigman, M. and Vinckier, F. 2005. "The Neural Code for Written Words: A Proposal." *Trends in Cognitive Sciences*, 9(7): 335-341.
- Fox, S.I. 2001. *Human Physiology*. Seventh Edition. McGraw-Hill Education: New York, New York.
- Desobry, F., Davy, M., and Doncarli, C. 2005. "An Online Kernel Change Detection Algorithm." *IEEE Transactions on Signal Processing*, 8: 2961–2974.
- Eisenberger, N.I., Lieberman, M.D. and Williams, K.D., 2003. "Does Rejection Hurt? An fMRI Study of Social Exclusion." *Science*, 302(5643): 290-292.
- Ekman, P., 1973. "Cross-cultural Studies of Facial Expression." *Darwin and Facial Expression: A Century of Research in Review*: 169-222.

- Ekman, P., 1999. "Basic Emotions." *Handbook of Cognition and Emotion*, 98: 45-60.
- Ekman, P. and Friesen, W.V., 1976. "Measuring Facial Movement." *Environmental Psychology and Nonverbal Behavior*, 1(1): 56-75.
- Ekman, P. and Cordaro, D., 2011. "What is Meant by Calling Emotions Basic." *Emotion Review*, 3(4): 364-370.
- Engelbert, M., & Carruthers, P. 2010. "Introspection." *Wiley Interdisciplinary Reviews: Cognitive Science*, 1: 245-253.
- Epstein, R.A., 2008. "Parahippocampal and Retrosplenial Contributions to Human Spatial Navigation." *Trends in cognitive sciences*, 12(10): 388-396.
- Feinstein, J.S., Adolphs, R., Damasio, A. and Tranel, D., 2011. "The Human Amygdala and the Induction and Experience of Fear." *Current Biology*, 21(1): 34-38.
- Figdor, C. 2010. "Neuroscience and the Multiple Realization of Cognitive Functions." *Philosophy of Science* 77(3): 419-456.
- Figdor, C. 2011. "Semantics and Metaphysics in Informatics: Toward an Ontology of Tasks (a Reply to Lenartowicz et al. 2010, Towards an Ontology of Cognitive Control)." *Topics in Cognitive Science*, 3(2): 222-226.
- Fodor, J. 1999. "Diary: Why the Brain?" *London Review of Books*, 19(30).
- Fox, M. D., and Raichle, M. E. 2007. "Spontaneous Fluctuations in Brain Activity Observed with Functional Magnetic Resonance Imaging." *Nature Reviews Neuroscience*, 8(9): 700-711.
- Fox, M. D., Snyder, A.Z., Vincent, J. L., Corbetta, M., Van Essen, D. C. & Raichle, M. E. 2005. "The Human Brain is Intrinsically Organized into Dynamic, Anticorrelated Functional Networks." *Proceedings of the National Academy of Sciences* 102 (27): 9673–9678.
- Fox, S.I. 2001. *Human Physiology*. Seventh Edition. McGraw-Hill Education: New York, New York.
- Frank, M.J. and Badre, D., 2015. "How Cognitive Theory Guides Neuroscience." *Cognition*, 135: 14-20.
- Franklin, L. R. 2005. "Exploratory Experiments." *Philosophy of Science*, 72(5): 888-899.
- Friston, K. J. 2011. "Functional and Effective Connectivity: a Review." *Brain Connectivity*, 1(1): 13-36.



- Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ. 1997. "Psychophysiological and Modulatory Interactions in Neuroimaging." *Neuroimage*, 6: 218–229.
- Garson, J. 2011. "Selected Effects and Causal Role Functions in the Brain: The Case for an Etiological Approach to Cognitive Neuroscience." *Biology and Philosophy*, 26: 547-565.
- Gauthier, I. and Tarr, M.J. 1997. "Becoming a "Greeble" Expert: Exploring Mechanisms for Face Recognition." *Vision Research*, 37(12): 1673-1682.
- Gauthier, I., Skudlarski, P., Gore, J.C. and Anderson, A.W., 2000. "Expertise for Cars and Birds Recruits Brain Areas Involved in Face Recognition." *Nature Neuroscience*, 3(2): 191-197.
- Getting, Peter A. 1989. "Emerging Principles Governing the Operation of Neural Networks." *Annual Review of Neuroscience*, 12: 185-204.
- Glover, G. H., & Lee, A. T. 1995. "Motion Artifacts in fMRI: Comparison of 2DFT with PR and Spiral Scan Methods." *Magnetic Resonance in Medicine*, 33(5): 624-635.
- Greene, J. and Haidt, J., 2002. "How (and Where) Does Moral Judgment Work?." *Trends in Cognitive Sciences*, 6(12): 517-523.
- Gregory, A. W., Nason, J. M., & Watt, D. G. 1996. "Testing for Structural Breaks in Cointegrated Relationships." *Journal of Econometrics*, 71: 321-341.
- Greicius, M. D., Kiviniemi, V., Tervonen, O., Vainionpää, V., Alahuhta, S., Reiss, A. L., & Menon, V. 2008. "Persistent Default-mode Network Connectivity during Light Sedation." *Human Brain Mapping*, 29(7): 839-847.
- Greicius, M. D., Krasnow, B., Reiss, A. L. & Menon, V. 2003. "Functional Connectivity in the Resting Brain: a Network Analysis of the Default Mode Hypothesis." *Proceedings of the National Academy of Sciences*, 100 (1): 253-258.
- Grill-Spector, K., Sayres, R., and Ress, D. 2006. "High-Resolution Imaging Reveals Highly Selective Nonface Clusters in the Fusiform Face Area." *Nature Neuroscience*, 9: 1177-1185.
- Gusnard, D. A., Akbudak, E. Shulman, G. L. & Raichle, M. E. 2001. "Medial Prefrontal Cortex and Self-referential Mental activity: Relation to a Default Mode of Brain Function." *Proceedings of the National Academy of Sciences*, 98 (7): 4259–4264.
- Gusnard, D. A., & Raichle, M. E. 2001. "Searching for a Baseline: Functional Imaging and the Resting Human Brain." *Nature Reviews Neuroscience*, 2(10): 685–694.
- Hamann, S., 2012. "Mapping Discrete and Dimensional Emotions onto the Brain: Controversies and Consensus." *Trends in Cognitive Sciences*, 16(9): 458-466.

- Hatfield, G., 2000. "The Brain's 'New' Science: Psychology, Neurophysiology, and Constraint." *Philosophy of Science*, 67: S388-S403.
- Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L. and Pietrini, P., 2001. "Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex." *Science*, 293(5539): 2425-2430.
- He, B. J., Snyder, A. Z., Zempel, J. M., Smyth, M. D., & Raichle, M. E. 2008. "Electrophysiological Correlates of the Brain's Intrinsic Large-scale Functional Architecture." *Proceedings of the National Academy of Sciences*, 105(41): 16039-16044.
- Heiser, M., Iacoboni, M., Maeda, F., Markus, J. and Mazziotta, J.C. 2003. "The Essential Role of Broca's Area in Imitation." *European Journal of Neuroscience*, 17: 1123-1128.
- Hodgkin, J. 1998. "Seven Types of Pleiotropy." *International Journal of Developmental Biology*, 42: 501-505.
- Henson, R.A. 2005. "What Can Functional Neuroimaging Tell the Experimental Psychologist?", *Quarterly Journal of Experimental Psychology*, 58A (2): 193–233.
- Henson, R., 2006. "Forward Inference Using Functional Neuroimaging: Dissociations versus Associations." *Trends in Cognitive Sciences*, 10(2): 64-69.
- Henson, R.N., Rugg, M.D., Shallice, T., Josephs, O. and Dolan, R.J., 1999. "Recollection and Familiarity in Recognition Memory: an Event-related Functional Magnetic Resonance Imaging Study." *The Journal of Neuroscience*, 19(10): 962-3972.
- Hohwy, J. 2013. *The Predictive Mind*. Oxford University Press.
- Ibañez, A., Gleichgerricht, E., and Manes, F. 2010. "Clinical Effects of Insular Damage in Humans." *Brain Structure and Function*, 214(5-6): 397-410.
- Izard, C.E., 2011. "Forms and Functions of Emotions: Matters of Emotion–cognition Interactions." *Emotion Review*, 3(4): 371-378.
- Johnston, J. M., Vaishnavi, S. N., Smyth, M. D., Zhang, D., He, B. J., Zempel, J. M., Shimony, J. S., Snyder, A. Z., & Raichle, M. E. 2008. "Loss of Resting Interhemispheric Functional Connectivity after Complete Section of the Corpus Callosum." *The Journal of Neuroscience*, 28(25): 6453-6458.
- Jones, E. E., & Harris, V. A. 1967. "The Attribution of Attitudes." *Journal of Experimental Social Psychology*, 3: 1–24.
- Kanwisher, N., 2010. "Functional Specificity in the Human Brain: a Window into the Functional Architecture of the Mind." *Proceedings of the National Academy of Sciences*, 107(25): 11163-11170.

- Kaplan, J.T., Man, K. and Greening, S.G., 2015. "Multivariate Cross-classification: Applying Machine Learning Techniques to Characterize Abstraction in Neural Representations." *Frontiers in Human Neuroscience*, 9(151): 1-12.
- Khalidi, M.A., 2013. *Natural Categories and Human Kinds: Classification in the Natural and Social Sciences*. Cambridge University Press.
- Khila, A., Abouheif, E., and Rowe, L. 2014. "Comparative Functional Analyses of *Ultrabithorax* Reveals Multiple Steps and Paths to Diversification of Legs in the Adaptive Radiation of Semi-Aquatic Insects." *Evolution*, 68(8): 2159-2170.
- Kipps, C.M., Duggins, A.J., McCusker, E.A. and Calder, A.J., 2007. "Disgust and Happiness Recognition Correlate with Anteroventral Insula and Amygdala Volume Respectively in Preclinical Huntington's Disease." *Journal of Cognitive Neuroscience*, 19(7): 1206-1217.
- Klein, C., 2010a. "Images are Not the Evidence in Neuroimaging." *The British Journal for the Philosophy of Science*, 61(2): 265-278.
- Klein, C., 2010b. "Philosophical Issues in Neuroimaging." *Philosophy Compass*, 5(2): 186-198.
- Klein, C. 2012. "Cognitive Ontology and Region-versus Network-Oriented Analyses." *Philosophy of Science*, 79(5): 952-960.
- Klein, C. 2014. "The Brain at Rest: What it is Doing and Why that Matters." *Philosophy of Science*, 81: 974–985.
- Laird, A.R., Lancaster, J.J. and Fox, P.T., 2005. "Brainmap." *Neuroinformatics*, 3(1), pp.65-77.
- Lashley, K., 1933 Integrative Functions of The Cerebral Cortex. *Physiological Review*, 13: 1-42.
- Laumann, T. O., Gordon, E. M., Adeyemo, B., Snyder, A. Z., Joo, S. J., Chen, M. Y., Gilmore, A. W., McDermott, K. B., Nelson, S. M., Dosenbach, N. U. F., Schlaggar, B. L., Mumford, J. A., Poldrack, R. A., & Petersen, S. E. (2015). Functional System and Areal Organization of a Highly Sampled Individual Human Brain. *Neuron*, 87(3): 657-670.
- Lenartowicz, A., Kalar, D. J., & Congdon, E. 2010. "Towards an Ontology of Cognitive Control." *Topics in Cognitive Science*, 2(4): 678-692.
- Leutgeb, S., Leutgeb, J.K., Barnes, C.A., Moser, E.I., McNaughton, B.L., and Moser, M.B. 2005. "Independent Codes for Spatial and Episodic Memory in Hippocampal Neural Ensembles." *Science*, 309(5734): 619-623.
- Liljeholm, M. and O'Doherty, J.P. 2012. "Contributions of the Striatum to Learning, Motivation, and Performance: An Associative Account." *Trends in Cognitive Sciences*, 16(9): 467-475.

- Lindquist, K.A. and Barrett, L.F., 2008. "Constructing Emotion the Experience of Fear as a Conceptual Act." *Psychological Science*, 19(9): 898-903.
- Lindquist, K.A., Wager, T.D., Kober, H., Bliss-Moreau, E. and Barrett, L.F., 2012. "The Brain Basis of Emotion: a Meta-analytic Review." *Behavioral and Brain Sciences*, 35(03): 121-143.
- Liversedge, S.P. and Findlay, J.M. 2000. "Saccadic Eye Movement and Cognition." *Trends in Cognitive Sciences*, 4(1): 6-14.
- Lloyd, D. 2000. "Terra Cognita: From Functional Neuroimaging to the Map of the Mind." *Brain and Mind*, 1(1): 93-116.
- Logothetis, N.K., 2003. "The Underpinnings of the BOLD Functional Magnetic Resonance Imaging Signal." *The Journal of Neuroscience*, 23(10): 3963-3971.
- Logothetis, N. K. 2008. "What we Can Do and what We Cannot Do with fMRI." *Nature*, 453(7197): 869-878.
- Lowe, M. J., Mock, B. J., & Sorenson, J. A. 1998. "Functional Connectivity in Single and Multislice Echoplanar Imaging using Resting-state Fluctuations." *Neuroimage*, 7(2): 119-132.
- MacDonald, G. and Leary, M.R., 2005. "Why Does Social Exclusion Hurt? The Relationship between Social and Physical Pain." *Psychological bulletin*, 131(2): 202.
- Machamer, P., Darden, L., and Craver, C. 2000. "Thinking about Mechanisms." *Philosophy of Science*, 67(1): 1-25.
- Machery, E. 2012. "Dissociations in Neuropsychology and Cognitive Neuroscience." *Philosophy of Science*, 79 (4): 490-518.
- Machery, E. 2013. "In Defense of Reverse Inference." *The British Journal for the Philosophy of Science*, 65(2): 251-267.
- McIntosh, A.R. 2000. "Toward a Network Theory of Cognition." *Neural Networks*, 13: 861-876.
- Meehl, P.E., 1978. "Theoretical Risks and Tabular Asterisks: Sir Karl, Sir Ronald, and the Slow Progress of Soft Psychology." *Journal of Consulting and Clinical Psychology*, 46(4): 806.
- Menon, V. and Uddin, L.Q. 2010. "Saliency, Switching, Attention and Control: A Network Model of Insula Function." *Brain Structure and Function*, 214 (5-6): 655-667.

- Mesulam, M-M. 1990. "Large-Scale Neurocognitive Networks and Distributed Processing for Attention, Language, and Memory." *Annals of Neurology*, 28: 597-613.
- Mikhail, J., 2007. "Universal moral grammar: Theory, Evidence and the Future." *Trends in Cognitive Sciences*, 11(4): 143-152.
- Milner, B., Corkin, S. and Teuber, H.L., 1968. "Further Analysis of the Hippocampal Amnesic Syndrome: 14-year Follow-up Study of HM." *Neuropsychologia*, 6(3): 215-234.
- Mitchell, J.P. 2008. "Activity in Right Temporo-Parietal Junction is not Selective for Theory of Mind." *Cerebral Cortex*, 18(2): 262-271.
- Morcom, A. M. & Fletcher, P. C. 2007. "Does the Brain have a Baseline? Why we Should be Resisting a Rest." *Neuroimage* 37 (4): 1073–1082.
- Mundale, J., 2002. "Concepts of Localization: Balkanization in the Brain." *Brain and Mind*, 3(3): 313-330.
- Nathan, M. J., and Del Pinal, G. 2016. "Mapping the Mind: Bridge Laws and the Psycho-neural Interface." *Synthese*, 193(2): 637-657.
- Neander, K. 1991. "Functions as Selected Effects: The Conceptual Analyst's Defense." *Philosophy of Science*, 58(2): 168-184.
- Nisbett, R. E., & Ross, L. (1991). *The Person and the Situation: Perspectives of Social Psychology*. New York: McGraw-Hill Publishing.
- Norman, K.A., Polyn, S.M., Detre, G.J. and Haxby, J.V., 2006. "Beyond Mind-reading: Multi-voxel Pattern Analysis of fMRI Data." *Trends in Cognitive Sciences*, 10(9): 424-430.
- Patterson, K. and Plaut, D.C., 2009. "Shallow Draughts Intoxicate the Brain": Lessons from Cognitive Science for Cognitive Neuropsychology. *Topics in Cognitive Science*, 1(1): 39-58.
- Pelphrey, K.A., Morris, J.P. and McCarthy, G., 2004. "Grasping the Intentions of Others: the Perceived Intentionality of an Action Influences Activity in the Superior Temporal Sulcus during Social Perception." *Journal of Cognitive Neuroscience*, 16(10): 1706-1716.
- Petersen, S.E. and Fiez, J.A. 1993. "The Processing of Single Words Studied With Positron Emission Tomography." *Annual Review of Neuroscience*, 16: 509-530.
- Pinel, P., Piazza, M., Le Bihan, D., Dehaene, S. 2004. "Distributed and Overlapping Cerebral Representations of Number, Size, and Luminance During Comparative Judgments." *Neuron*, 41(6): 983-993.

- Plaut, D.C., 1995. "Double Dissociation without Modularity: Evidence from Connectionist Neuropsychology." *Journal of Clinical and Experimental Neuropsychology*, 17(2): 291-321.
- Poldrack, R. A. 2006. "Can Cognitive Processes be Inferred from Neuroimaging Data?" *Trends in Cognitive Sciences*, 10(2): 59-63.
- Poldrack, R.A., 2008. "The Role of fMRI in Cognitive Neuroscience: Where do we Stand?." *Current Opinion in Neurobiology*, 18(2): 223-227.
- Poldrack, R. A. 2010. "Mapping Mental Function to Brain Structure: How Can Cognitive Neuroimaging Succeed?." *Perspectives on Psychological Science*, 5(6): 753-761.
- Poldrack, R.A., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y., Parker, D.S., Sabb, F.W. and Bilder, R.M., 2011. "The Cognitive Atlas: Toward a Knowledge Foundation for Cognitive Neuroscience." *Frontiers in Neuroinformatics*, 5(0).
- Polger, T.W. and Shapiro, L.A., 2016. *The Multiple Realization Book*. Oxford University Press.
- Posner, M.I. and Raichle, M.E., 1998. "The Neuroimaging of Human Brain Function." *Proceedings of the National Academy of Sciences*, 95(3): 763-764.
- Povich, M., 2015. "Mechanisms and Model-Based Functional Magnetic Resonance Imaging." *Philosophy of Science*, 82(5): 1035-1046.
- Price, C.J. and Friston, K.J., 2002. "Degeneracy and Cognitive Anatomy." *Trends in Cognitive Sciences*, 6(10): 416-421.
- Price, C. J. and Friston, K. J. 2005. "Functional Ontologies for Cognition: The Systematic Definition of Structure and Function." *Cognitive Neuropsychology*, 22(3-4): 262-275.
- Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., Vogel, A. C., Laumann, T. O., Miezin, F. M., Schlaggar, B. L., & Petersen, S. E. (2011). Functional Network Organization of the Human Brain. *Neuron*, 72(4): 665-678.
- Power, J. D., Schlaggar, B. L., & Petersen, S. E. (2014). Studying Brain Organization via Spontaneous fMRI Signal. *Neuron*, 84(4): 681-696.
- Purdon, P. L., & Weisskoff, R. M. (1998). Effect of Temporal Autocorrelation due to Physiological Noise and Stimulus Paradigm on Voxel-level False-positive Rates in fMRI. *Human Brain Mapping*, 6(4): 239-249.
- Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A Default Mode of Brain Function. *Proceedings of the National Academy of Sciences*, 98 (2): 676-682.

- Raichle, M. E. (2009). A Paradigm Shift in Functional Brain Imaging. *The Journal of Neuroscience*, 29(41): 12729-12734.
- Rathkopf, C. A. 2013. "Localization and Intrinsic Function." *Philosophy of Science*, 80(1): 1-21.
- Redner, R. A., & Walker, H. F. (1984). Mixture Densities, Maximum Likelihood, and the EM Algorithm. *SIAM Review*, 26: 195-239.
- Rosazza, C., & Minati, L. (2011). Resting-state Brain Networks: Literature Review and Clinical Applications. *Neurological Sciences*, 32(5): 773-785.
- Roskies, A. L. 2009. "Brain-Mind and Structure-Function Relationships: A Methodological Response to Coltheart." *Philosophy of Science*, 76(5): 927-939.
- Roskies, A.L., 2010. "Saving subtraction: A reply to Van Orden and Paap." *The British Journal for the Philosophy of Science*, 61(3): 635-665.
- Ryle, G., 1945. "Knowing How and Knowing That: The Presidential Address." In *Proceedings of the Aristotelian society*, 46: 1-16. Aristotelian Society, Wiley.
- Sabb, F.W., Bearden, C.E., Glahn, D.C., Parker, D.S., Freimer, N. and Bilder, R.M., 2008. "A Collaborative Knowledge base for Cognitive Phenomics." *Molecular Psychiatry*, 13(4): 350-360.
- Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I.P., Lampinen, J., Vuilleumier, P., Hari, R., Sams, M. and Nummenmaa, L., 2015. "Discrete Neural Signatures of Basic Emotions." *Cerebral Cortex*, p.bhv086.
- Saxe, R. and Kanwisher, N. 2003. "People Thinking about People: fMRI Investigations of Theory of Mind. *Neuroimage*, 19: 1835-1842.
- Saxe, R., Brett, M., & Kanwisher, N. (2006). Divide and Conquer: a Defense of Functional Localizers. *Neuroimage*, 30(4): 1088-1096.
- Scarantino, A., 2012. "Functional Specialization does not Require a One-to-one Mapping between Brain Regions and Emotions." *Behavioral and Brain Sciences*, 35(03): 161-162.
- Scarantino, A. and Griffiths, P., 2011. "Don't Give up on Basic Emotions." *Emotion Review*, 3(4): 444-454.
- Scargle, J. D. (1998). Studies in Astronomical Time Series Analysis. V. Bayesian Blocks, a New Method to Analyze Structure in Photon Counting Data. *The Astrophysical Journal*, 504: 405-418.
- Schmahmann, J. D. and Caplan, D. 2006. "Cognition, Emotion and the Cerebellum." *Brain*, 129(2): 290-292.

- Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E.N. & Saxe, R. 2009. "Distinct Regions of Right Temporo-Parietal Junction are Selective for Theory of Mind and Exogenous Attention." *PLoS One* 4: 1-7.
- Scoville, W.B. and Milner, B., 1957. "Loss of Recent Memory after Bilateral Hippocampal Lesions." *Journal of Neurology, Neurosurgery, and Psychiatry*, 20(1): 11.
- Schwitzgebel, E. 2008. "The Unreliability of Naive Introspection." *Philosophical Review*, 117: 245-273.
- Seth, A. K., Chorley, P., & Barnett, L. C. 2013. "Granger Causality Analysis of fMRI BOLD Signals is Invariant to Hemodynamic Convolution but not Downsampling." *NeuroImage*, 65: 540-555.
- Shallice, T., 1988. *From Neuropsychology to Mental Structure*. Cambridge University Press.
- Smith, G.T., 2005. "On construct Validity: Issues of Method and Measurement." *Psychological Assessment*, 17(4): 396.
- Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., Filippini, N., Watkins, K. E., Toro, R., Laird, A. R., & Beckmann, C. F. 2009. "Correspondence of the Brain's Functional Architecture During Activation and Rest." *Proceedings of the National Academy of Sciences*, 106(31): 13040-13045.
- Snyder, A. Z. & Raichle, M. E. 2012. "A Brief History of the Resting State: the Washington University Perspective." *Neuroimage*, 62(2): 902-910.
- Sporns, O., 2011. *Networks of the Brain*. MIT press.
- Steinle, F. 1997. "Entering New Fields: Exploratory uses of Experimentation." *Philosophy of Science*, 64: S65-S74.
- Sullivan, Jacqueline A. 2009. "The Multiplicity of Experimental Protocols: a Challenge to Reductionist and Non-reductionist Models of the Unity of Neuroscience." *Synthese* 167, 3: 511-539.
- Sullivan, Jacqueline A. 2010. "Reconsidering 'Spatial Memory' and the Morris Water Maze." *Synthese* 177, 2: 261-283.
- Suri, R.E. and Schultz, W. 2001. "Temporal Difference Model Reproduces Anticipatory Neural Activity." *Neural Computation*, 13: 841-862.
- Swick, D., Ashley, V. and Turken, U., 2011. "Are the Neural Correlates of Stopping and not going Identical? Quantitative Meta-analysis of Two Response Inhibition Tasks." *Neuroimage*, 56(3): 1655-1665.



- Tettamanti, M. and Weniger, D. 2006. "Broca's Area: A Supramodal Hierarchical Processor?" *Cortex*, 42: 491-494.
- Tettamanti, M., Rotondi, I., Perani, D., Scotti, G., Fazio, F., Cappa, S.F., and Moro, M. "Syntax Without Language: Neurobiological Evidence for Cross-Domain Syntactic Computations." *Cortex* 45(7): 825-838.
- Thagard, P., 1990. "Concepts and Conceptual Change." *Synthese*, 82(2): 255-274.
- Thagard, P., 2012. *The Cognitive Science of Science: Explanation, Discovery, and Conceptual Change*. MIT Press.
- Uttal, W. R. 2001. *The New Phrenology: The Limits of Localizing Cognitive Processes in the Brain*. The MIT press.
- Uttal, W. R. 2008. *Distributed Neural Systems: Beyond the New Phrenology*. Sloan Pub.
- Van Orden, G. C. and Paap, K.R. 1997. "Functional Neuroimages Fail to Discover Pieces of Mind in the Parts of the Brain." *Philosophy of Science*, 64: S85-S94.
- Vemuri, K., and Surampudi, B. R. 2015. "Evidence of Stimulus Correlated Empathy Modes—Group ICA of fMRI data." *Brain and Cognition*, 94: 32-43.
- Vigouroux, R., 1879. "Sur le Role De la Resistance Electrique des Tissus Dans L'Electro-diagnostic". *Comptes Rendus Société de Biologie*, 31: 336-339.
- Vincent, J. L., Patel, G. H., Fox, M. D., Snyder, A. Z., Baker, J. T., Van Essen, D. C., Zempel, J. M., Snyder, L. H., Corbetta, M., & Raichle, M. E. 2007. "Intrinsic Functional Architecture in the Anaesthetized Monkey Brain." *Nature*, 447(7140): 83-86.
- Wager, T. D., and Barrett, L.F. 2004. "From Affect to Control: Functional Specialization of the Insula in Motivation and Regulation." Online Publication at <http://www.apa.org/psycextra>.
- Wang, G. Tanifuji, M. and Tanaka, K. 1998. "Functional Architecture in Monkey Inferotemporal Cortex Revealed by In Vivo Optical Imaging." *Neuroscience Research*, 32: 33-46.
- White, B. R., Bauer, A. Q., Snyder, A. Z., Schlaggar, B. L., Lee, J. M., & Culver, J. P. 2011. "Imaging of Functional Connectivity in the Mouse Brain." *PloS one*, 6(1): e16322.
- Woo, C.W., Koban, L., Kross, E. Lindquist, M.A., Banich, M.T., Ruzic, L., Andrews-Hanna, J.R. and Wager, T.D. 2014. "Separate Neural Representations for Physical Pain and Social Rejection." *Nature Communications*, 5.
- Wouters, A. 2005. "The Function Debate in Philosophy." *Acta Biotheoretica*, 53: 123-151.

- Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. 2011. "The Organization of the Human Cerebral Cortex Estimated by Intrinsic Functional Connectivity." *Journal of Neurophysiology*, 106(3): 1125-1165.
- Zheng, J., & Frey, H. C. 2004. "Quantification of Variability and Uncertainty using Mixture Distributions: Evaluation of Sample Size, Mixing Weights, and Separation between Components." *Risk Analysis*, 24(3): 553-571.